FINAL REPORT

Application of Hyperspectral Techniques to Monitoring and Management of Invasive Plant Species

SERDP Project SI-1143

University of California at Davis
Dr. Susan L. Ustin
Mary Andrews
Margaret E. Andrew
Yen-Ben Cheng
Deanne DiPietro
John Kefauver
Shawn Kefauver
Nina Noujdina
Karen S. Olmstead
Carlos M. Ramirez
Edward T. Tom
Emma Underwood

Dynamac Corporation
Teresa K. Magee
Lynne S. McAllister
Ted L. Ernst

SERDP

Strategic Environmental Research and Development Program

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to the Department of Defense, Executive Services and Communications Directorate (0704-0188). Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE *(DD-MM-YYYY)* 09/01/2008	2. REPORT TYPE Final Technical Report	3. DATES COVERED *(From - To)* 7/2000 - 12/2006

4. TITLE AND SUBTITLE

Application of Hyperspectal Techniques to Monitoring & Management of Invasive Plant Species Infestation

5a. CONTRACT NUMBER

DACA72-00-C-0012

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Ustin, Susan L.
Olmstead, Karen S.
Noujdina, Nina
Cheng,Yen-Ben

5d. PROJECT NUMBER

SI-1143

5e. TASK NUMBER

13.1

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of California, Davis/ CalSpace Center of Excellence
The Barn,One Shields Avenue
Davis, CA 95616

8. PERFORMING ORGANIZATION REPORT NUMBER

N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Strategic Environmental Research & Development Program (SERDP)
901 North Stuart St. Suite 303
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)

N/A

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

N/A

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This report summarizes and integrates the main findings, using several case studies to illustrate conclusions. Case studies include cheatgrass and Russian knapweed at Yakima Training Center, showing that multiple dates that combine different growing seasons improve map accuracy. Also included are maps for kudzu and Johnson lovegrass at Fort Benning, tamarisk at Yuma Proving Ground and phragmites at Aberdeen Proving Ground. Examples were chosen to illustrate a range of problems in correctly mapping invasive plant species under different environmental conditions. A detailed analysis of the CART results for Vandenberg Air Force Base show how it can be used to predict likely locations for spread and the most probable causes for invasion at specific locations. A CART analysis for kudzu at Ft. Benning is included and the same method was applied to tree decline at Ft. Benning. A User's Guide is provided in the appendix as is a hyperspectral training tutorial.

15. SUBJECT TERMS

cheatgrass, Russian Knapweed, kudzu, Johnson lovegrass, tamarisk, phragmites, Yakima Training Center, Fort Benning, Vandenberg Air Force base, Yuma Proving Ground, Aberdeen Proving Ground, hyperspectral mapping, CART model

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE	UU	321	Susan L. Ustin
U	U	U			19b. TELEPHONE NUMBER *(Include area code)* 530-752-0621

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Project SI-1143: Application of Hyperspectral Techniques to Monitoring and Management of Invasive Plant Species Infestation

Final Report

Strategic Environmental Research and Development Program

SERDP/ESTCP
901 N. Stuart Street
Suite 303
Arlington, VA 22203-1821

Attention:
Dr. John Hall
Mr. Johnathan Thigpen

Submitted on Behalf of the Regents of the University of California by:

Dr. Susan L. Ustin (Principal Investigator)
Mary Andrews
Margaret E. Andrew
Yen-Ben Cheng
Deanne DiPietro
John Kefauver
Shawn Kefauver
Nina Noujdina
Karen S. Olmstead
Carlos M. Ramirez
Edward T. Tom
Emma Underwood
University of California at Davis

Teresa K. Magee
Lynne S. McAllister
Ted L. Ernst
Dynamac Corporation

January 2008

Table of Contents

List of Figures

List of Tables

List of Acronyms

1. Image Data

AVHRR	Advanced Very High Resolution Radiometer
AVIRIS	Advanced Visible Infrared Imaging Spectrometer
ETM or ETM+	Landsat Enhanced Thematic Mapper
HSI	Airborne Hyperspectral Imagery
LiDAR	Light Detection And Ranging
MODIS	MODerate Resolution Imaging Spectrometer
SPOT	Systeme pour l'Observation de la Terre
TM	Landsat Thematic Mapper

2. Agencies

AFB	Air Force Base
AFRL	Air Force Research Laboratory
NASA	National Aeronautics and Space Administration
NPS	National Park Service
SITAC	Spectral Information Technology Application Center
USGS	United States Geological Survey

3. University of California

CSTARS	Center for Spatial Technologies and Remote Sensing
UC Davis	University of California Davis

4. Image Analysis and Geographic Analysis

ACORN	Atmospheric CORrection Now
ATCOR	Atmospheric Correction and Haze Reduction algorithm
CART	Classification and Regression Tree
ENVI	Environment for Visualization of Images
FLAASH	Fast Line-of-sight Atmospheric Analysis of Spectral. Hypercubes
GIS	Geographic Information Systems
GPS	Geographic Positioning System
LPD	Loblolly Pine Decline

ML	Maximum Likelihood Classifier
MNF	Minimum Noise Fraction
MTMF	Mixture Tuned Matched Filter
NDVI	Normalized Difference Vegetation Index
VI(s)	Vegetation Index(es), general
Various Physiological Indexes Acronyms	See Tables 3 and 4 for Physiological Indexes Acronyms

5. Measurement Units

cm	Centimeter (0.1m)
km	Kilometers (1000 m)
M	Meters (1m)
µm	Micrometer (10^{-6}m; 0.000001 m)
n-dimensional space	*Multi*-dimensional space, where n = number dimensions (where n-spectral dimensions is number of spectral bands or band transforms, e.g., eigenvectors; i.e., a 224 spectral band data set has 224 spectral dimensions)
nm	Nanometer (10^{-9}m; 0.000000001 m)
pixel	Picture Element (smallest resolvable spatial unit in an image)

Acknowledgments

Executive Summary

This report summarizes and synthesizes the major findings and activities of the SERDP project "Application of Hyperspectral Techniques to Monitoring and Management of Invasive Plant Species Infestation" for the period from 2000 through 2006. The study explored the potential for hyperspectral remote sensing technologies (an imaging technology that measures many (100s) of contiguous narrow spectral bands across the visible to mid-infrared spectrum (400-2500 nanometers), especially those having the capability of high spatial resolution, to provide a basis to detect invasive plant species. If suitable, this technology could be used to improve mapping and monitoring for early detection and management of invasive plant species. This study examined the potential for detection of invasive plant species on six military bases (Aberdeen Proving Ground, Camp Pendleton, Fort Benning, Vandenberg Air Force Base, Yakima Training Center, and Yuma Proving Ground) using hyperspectral imagery. These military bases were located in different ecoregions of the continental United States allowing the project to test the detectability of invasive plant species that grow over a wide range of environmental and climate conditions. Additionally, the study selected invasive plant species that had different growth forms, from herbs to vines and shrubs, and native ecosystems ranging from aquatic, salt marsh, riparian zone, chaparral, warm and cold desert scrub, to deciduous forest.

The project used NASA's airborne Advanced Visible Infrared Imaging Spectrometer (AVIRIS) with 224 contiguous narrow spectral bands, flown at high spatial resolution (~ 9-12 m^2 pixels [picture elements]), over parts of each base. The airborne flightlines were oriented to include invaded and uninvaded native vegetation and generally, multiple flightlines were acquired at each base in order to include a range of conditions and growth stages to test detection of invasive plant species and native species. At two bases, flights were made in more than one season, providing a basis to comment on use of multiple-date data for invasive plant species detection.

Significant Findings and Lessons Learned

A number of significant findings and lessons learned were derived from this SERDP study. The findings of this study are consistent with other recent studies of invasive plant species and other vegetation mapping studies using hyperspectral data, indicating that these results have general applicability.

1. Portability of Methods. This study demonstrated that hyperspectral methodology has broad applicability for invasive plant species mapping at these six military installations and is extendable to other bases. The analysis of hyperspectral imagery was tailored to address unique spectral, spatial, and temporal characteristics of each invasive plant species studied. In each case these methods were able to identify the invasive plant species to satisfactorily high accuracy using field-acquired training data combined with standard processing methods and validated with independent field data.

2. Uniqueness of Invasive Plant Species within the Native Plant Community. The biology of the invasive plant species and the native plant communities must be considered in identifying the best choice for the spatial, spectral, and temporal resolutions. Detection depends on the presence of a structural and/or biochemical difference between invasive plant species and native species.

Detection is best when different growth forms are present between the invasive and native species, e.g., an invasive grass in a shrubland or vise versa. In cases where the growth forms of invasive and native species are similar, differences in size and density of invasive species vs. native contribute to detection. Timing data acquisition to periods when phenological differences (timing of seasonal growth) are greatest between native and invasive plant species also contributes significantly to detection. Phenological differences capture both structural (herbaceous, woody, leaf type; flowering, fruiting) and biochemical differences (concentrations of pigments, water, organic matter (e.g., woody stems) between species. Only hyperspectral imagery can capture differences in concentrations of plant pigments, canopy water, and canopy dry material and differences in soil properties and this information aids in detection. It is easier to differentiate structural differences (generally causing changes in light scattering) from biochemical differences (generally caused by absorption of energy at specific wavelengths) using hyperspectral imagery than multispectral imagery. The ability to differentiate species using hyperspectral imagery will continue to improve as the technology improves and as radiative transfer models improve. These are a class of physically based models that characterize the processes of light scattering and absorption that are due to leaf and plant canopy characteristics. As this understanding increases, the ability to discriminate species more precisely will improve.

3. Spatial Resolution. The best choice of spatial resolution requires some knowledge of the within class spatial (distributional) heterogeneity of the invasive plant species and the heterogeneity of the background vegetation classes. For the sites studied, patches of invasive plant species were routinely detected in hyperspectral imagery at spatial resolutions of 3-4 m x 3-4 m (9 to12 m^2). The actual pixel size measured depends on the instrument design and the height of the plane above the ground surface. Hyperspectral image data was not available at higher spatial resolution therefore no conclusions are implied from this study about whether improved maps would result from better resolution. However implications from geostatistical models like krigging, indicate that if spatial resolution is too fine (e.g., within canopy crown measurements) that the pixel-to-pixel variability may overwhelm correlation to spatial patterns at somewhat coarser scales. Similarly, if the pixel size is larger than the spatial resolution where correlations occur, then the data is not useful. Therefore, a geostatistical analysis of the spatial heterogeneity of the invasive species and the native plant background would provide the critical information needed to *a priori* determine the "best" spatial resolution.

Most cases found that spatial resolutions higher than 4m x 4m pixels were not required, if the analysis used the full spectrum from the visible (400 nm) to the full extent of the reflected solar infrared (~2500 nm). In this study's examination of spatial resolution, in at least one case, the heterogeneity of the background plant community was actually best mapped at a coarse spatial resolution of 30 m when the full hyperspectral spectral resolution (in these data, 224 spectral bands) was available. The improved mapping accuracy for invasive plant species in the larger pixels was attributed to reduced pixel-to-pixel heterogeneity of the background vegetation. In this study, identification of invasive species required densities of at least 20% of the total plant canopy. It is well established in the remote sensing literature that subpixel estimates of cover can be accurately estimated from hyperspectral data when sufficient spectral differences between class types (other plant species, soil, etc.) exist.

4. Use of Field Data. Field observations of patches of invasive plant species and non-invaded native communities should be located with Global Positioning System (GPS) points centered in homogeneous patches or as a polygon that maps the periphery of the patch. This method to obtaining training and validation data for the image analysis works well on small pixel sized data (e.g., the scale used in this study). Obtaining training data this way is simpler than making field spectral measurements of the range of dominant species in the habitat. This is partially due to the limited time per day that good field spectral measurements can be made (near solar noon and under clear sky conditions) and near the date of the overflight (typical rule of thumb is about 2 weeks on either side of the overflight). If field data are used to create training data for image analyses, it is critical that the field data and images be geolocated as precisely as possible to true locations so that GPS located field data are related to the correct pixels. The ability to geolocate pixels to correct geographic coordinates currently provides a practical limit of about one meter square spatial resolution. Nonetheless, given location errors it is difficult (except for very limited surveys) to geolocate pixels to base maps and to match field GPS data to pixels that are smaller than 2m x 2m. Thus, training data is best if mapped patches are several pixels in extent.

5. Requirements for Hyperspectral Data. In particular, this study found that although full spectrum measurements provided the best results when a biophysical difference (e.g., succulence vs. xeric foliage) was present between the invasive plant species and the native plant community, simpler methods like the band ratios that are used in spectral indexes were suitable for quick analysis. When the invasive plant species is present in dense and relatively homogeneous patches, all methods, including the simpler spectral index methods, were sufficient for identification. In no case did multispectral (4-6 band) analysis produce a better distribution map of invasive plant species than using the full resolution hyperspectral data.

5. Universality of Spectral Methods. In many cases, several methods using full spectrum techniques were found to be sufficient for mapping, including linear spectral mixing, spectral matching, and principal components analysis methods. The better detection derived from these methods is partially attributed to the smoothing that occurs across the full spectrum, relative to methods that use isolated bands. The shortwave infrared (1500-2500nm) significantly contributes to the detection of invasive plant species in a vegetated environment, despite the small absorption signals present in this part of the spectrum and the lower signal-to-noise of the detectors. Because multispectral instruments typically lack bands in the shortwave infrared, this is one reason why hyperspectral data produces better accuracy as is the redundancy that makes it possible to interpret an over-determined system.

6. Timing Image Acquisition. The spectral information in the data is enhanced if the timing of the data acquisition maximizes phenological differences between the invasive plant species and the native species. The goal in timing the collection of hyperspectral measurements is to capitalize on differences in plant cover and the fraction of bare soil and differences in the proportions of photosynthetic pigments, canopy water content, plant litter, and woody debris. Data collected in spring, when all species in a community tend to be actively growing, yields the poorest results for identifying invasive plant species in native communities. Periods of near-total dormancy for native and invasive species also produce poor results. Species differences are accentuated by acquiring the airborne hyperspectral data sometime *after* the peak of the growing season, and thus providing greater separation of the invasive plant species and native species.

Separation of invasive plant species from background vegetation is further enhanced using multiple-date data, particularly by combining data from different seasons, again emphasizing differences in timing of phenologic cycles between the invasive plant species versus the native vegetation. Declining costs and greater availability of airborne hyperspectral imagers by government and commercial vendors will make it feasible to time data collection to more optimal periods.

7. Minimum Detectability. This study was not able to detect invasive plant species in the presence of background vegetation when concentrations were below 20-40% (depending on the specific invasive plant specie and background). This implies that the earliest phase of invasive plant species establishment is not detectable or only marginally detectable. While this limitation may be improved by better quality higher spatial resolution data and/or improved analytical methods, it is likely that it cannot identify this first critical establishment stage from airborne data. Similarly, field observations are likely to miss this phase also since field sampling is severely limited in extent relative to the synoptic observations in images. It is likely that spatial modeling methods that predict risk of invasion and establishment of invasive species is the most approachable means for anticipating directions of spread once maps of current distributions of invasion are known.

8. Utilizing Spatial Information in Images. Most image processing packages focus on analysis of the spectral bands and consideration of the spatial context of the pixels is not part of the package of analytical tools. Recently, new programs e.g., eCognition, that consider the textural and spatial relationships have becoming more widely used. This spatial relationship information significantly adds to the spectral information in images. Methods that use spatial classifiers or merged spatial and spectral techniques are expected to continue to have increased applications. The hyperspectral datasets in this study were integrated into GIS databases with other site characteristics, e.g., soil, topography, and land use to both evaluate the invasive plant species identification and to develop an understanding of site conditions "at risk" for invasion. One type of spatial model, the Classification and Regression Tree (CART), was investigated in this study which showed that a GIS-based model can be developed to predict locations that exist on military installations that have characteristics that promote invasion by invasive plant species, thus implying that they are susceptible to invasion. There are a number of other spatial models that also are appropriate to be used to predict conditions favorable to invasion, e.g., genetic algorithm models (GAMs). Under the present limits of technology, a strategy focused on using the hyperspectral maps to identify where established patches of invasive plant species occur and then extending this to predict where propagules of invasive plant species might spread appears to be an achievable objective.

9. Integrating Imagery into a GIS Based Model Framework. Detection of invasive plant species from native species is best when the range of variability of other conditions in the environment is minimized. This can be done by eliminating all areas within the images that are irrelevant to the study (e.g., deleting open water pixels from further analysis when mapping terrestrial plant species). This is first accomplished by stratifying the scene into generalized components, identifying the native plant communities subject to invasion, and other conditions that are not suitable for invasion (e.g., impervious and developed areas, water bodies, etc.).

Pixels identified in non-relevant strata should be masked out (unused or deleted) from all further analyses, which will improve the accuracy of the final maps of invasive species.

10. Portability of Methods. The analytical tools used in this study were focused on image analysis methods available in commercial image processing packages because it was expected that these methods facilitate transfer of the technology to staff at military installations. Prior to the study, it was anticipated that some highly specialized spectral matching tools for image processing would be required but instead several methods commonly used in commercial hyperspectral image analysis programs, including linear mixture analysis, principal components analysis, and spectral fitting algorithms like "spectral angle mapper" and a method termed "mixture tuned matched filters" provided good results. It is thought that for many applications following a simple sequence of image analysis procedures, as described in the "user guide feature" later in this report, can be adopted successfully.

11. Technology Limitations and New Mapping Technologies. The technological limitations are rapidly changing and hyperspectral imagers are becoming available with significantly better spectral resolution, over expanded wavelength ranges (380nm to 3000nm), and with higher optical fidelity than the measurements used in this study. The next generation of high fidelity imaging spectrometers promise substantially better results than were obtained in this study (image data that was acquired from 2000-2004). These newer instruments have higher photon detection rates, therefore can measure smaller pixels if this is necessary (currently smaller than $1m^2$ pixels are possible). Improved geolocation capability, including ortho-rectification of images performed in flight is now achievable. Additionally, the software for correcting atmospheric processes using radiative transfer models to account for scattering and absorptions by gases and aerosols in the atmosphere are also improving and thus, will improve the quality of the reflectance calibration, enabling better detection and characterization of surface conditions.

During the period of this project, LiDAR data, airborne laser sensors that produce topographic maps of the ground elevation at 0.5m to 1m horizontal and a few centimeters of vertical resolution, and either the height of the canopy (a first and last return LiDAR) or a three-dimensional profile through the canopy (a full waveform LiDAR) have become and are becoming widely available. When combined with optical hyperspectral images, the merged data provide unprecedented information about the structure and biochemistry of species in the habitat, and the composition and density of the vegetation. These combined data have significant potential to revolutionize mapping species composition and detection of invasive species. It is likely that such data will improve the ability to measure invasions at an earlier stage of establishment than observed in this study.

Summary Table: Study sites, target invasive plant species, associated habitat, detection results, relative success and findings

Study Sites and Habitat	Target Invasive Plant Species	Best Analytical Method(s)	Detection/Results Success/Findings
Vandenberg Air Force Base Coastal Dunes	Iceplant - *Carpobrotus* spp. (*C. edulis*, *C. chilensis*, hybrids)	Maximum Likelihood on Minimum Noise Fraction	Pixel sized patches identified; high accuracy
Vandenberg Air Force Base Chaparral	Jubata Grass	Maximum Likelihood on Minimum Noise Fraction	Accuracy depends on inter-pixel variability in background vegetation
Camp Pendleton Marine Corp Base Grassland Riparian	Fennel - *Foeniculum vulgare* Giant reed - *Arundo donax*	Mixture Tuned Matched Filter Maximum Likelihood on Minimum Noise Fraction	Fennel detected even when grasslands dry; Giant reed sometimes confused with other riparian species.
Fort Benning Deciduous Forest	Lovegrass - *Eragrostis* spp. Kudzu – *Pueraria lobata*	Spectral Angle Mapper on Minimum Noise Fraction	Lovegrass not observed under tree canopies but detected along roads; kudzu was mapped in some locations aided by texture characteristics
Yuma Proving Ground Sonoran Desert	Mediterranean grass - *Schismus* spp. Asian mustard *Brassica tournefortii* Tamarisk - *Tamarix* spp. (*T. chinesis, T. aphylla*)	*Schismus* not measured due to drought; Any method for Tamarisk adjacent to irrigated agriculture; only Mixture Tuned Matched in dry washes	Schismus and *Brassica* not measured due to drought conditions at time of overflight; Most any method worked for Tamarisk receiving agricultural water in native landscape but near undetectable in dry washes

Yakima Training Center Great Basin Sagebrush	Cheatgrass – *Bromus tectorum* Russian knapweed-*Acroptilon repens*	Mixture Tuned Matched Filter	Mapping improved using combined spring and summer dates that takes advantage of phenological separation
Aberdeen Proving Ground Riparian Aquatic	Common reed – *Phragmites australis* Hydrilla – *Hydrilla verticillata*	In dense patches, Minimum Noise Fraction and physiological indexes; in mixed patches Mixture Tuned Matched Filters	*Phragmites* detection successful in environments with simpler structure than more heterogeneous landscape; unable to map *Hydrilla* because of too few training data and specular reflection off water.

1. Introduction

The rapid spread of invasive plant species is causing and has caused irreparable damage to the natural resources on military installations. The problem of invasive plant species has accelerated in recent years and is recognized as one of the most serious and rapidly increasing ecological threats faced by military environmental managers. Controlling and managing invasive plant species is a global environmental problem and is not unique to military installations. However, the significant role that ecosystem disturbance plays in facilitating the spread of noxious invasive plant species creates a challenge for military resource managers who must still fulfill the military mission while reducing or mitigating impacts on ecosystem structure and function. Troop, vehicle, and other operational and training activities on military bases cause the types of landscape disturbances that facilitate the spread of noxious invasive plant species. These problems are exacerbated by the inadvertent transmittal of invasive pests during military transport of people and equipment around the world. DOD, along with other federal agencies, has been mandated to develop proactive management that protects native species and preserves the structure and function of natural ecosystems.

2. Project Objectives

This study investigated the use of hyperspectral imagery, also termed imaging spectroscopy, to map invasive plant species in the presence of native vegetation. These remote sensing imagers, currently available on airborne and spaceborne platforms, have many narrow, contiguous spectral bands, producing a laboratory-like spectrum for each pixel in the image. This project addressed two of the objectives listed in the SERDP Statement of Need (SON): 1) develop methods to inventory the spatial extent, and where applicable, the densities of established populations of noxious weeds; and 2) develop methods to estimate the likelihood of noxious weed expansions into and/or invasion of new habitats/areas on military bases. To test the applicability and portability of image analysis methods under different environmental conditions, the study consisted of mapping 12 invasive plant species on six military bases. All airborne hyperspectral data for this project was provided by purchase order from the National Aeronautics and Space Administration (NASA) instrument, the Advanced Visible Infrared Imaging Spectrometer (AVIRIS) which measures 224 bands of approximately 10nm bandwidth over the spectral range from 400 to 2500 nm. This covers the visible (400-700nm) and reflected solar infrared (700-2500 nm) spectrum. It was anticipated that the spectral resolution of this instrument, when flown at low altitude to produce images of high, 9-12 m^2 pixel spatial resolution, would provide sufficient spectral separability to distinguish homogeneous patches of the target invasive plant species.

3. Background

3.1. Mapping Invasive plant species with Hyperspectral Imagery

The principal objective of the project was to develop and test a hyperspectral methodology for mapping invasive plant species. If successful, this technology could provide better and more accurate maps of invasive plant species on military installations that would allow base personnel to meet eradication and containment goals with more timely and cost effective management. A secondary goal was to develop methods with broad applicability to military installations that would also improve invasive plant species management at other bases. To provide the broadest basis for assessing transportability of the methods, the invasive plant species selected (Table 1) and native vegetation had a range of growth characteristics (herbaceous, grasses, shrubs, trees) chosen from military bases with different missions (Vandenberg AFB, Camp Pendleton, Yakima Training Center, Yuma Proving Ground, Ft. Benning, and Aberdeen Proving Ground), and located in different climate zones (Northwest Great Basin semi-desert, southern California Coast, Sonoran desert, Atlantic coast wetlands, and southeastern hardwoods). In each case, the base personnel identified 2-3 invasive plant species of greatest management concern. Their potential for identification using hyperspectral data was evaluated and species were selected that had some potential for detection, although not necessarily the easiest to detect. The goal was to test the methodology across a range of conditions that would help establish the general applicability of the technology. This set of species and bases allowed the project to compare invasive plant species and communities with different growth forms and climate conditions. The mix of invasive plant species and conditions at the different bases has allowed us to generalize results and compare methods. In this study, the unique spectral, spatial, and temporal characteristics of each invasive plant species was studied to assess the applicability of hyperspectral imaging (HSI) sensors to provide invasive species mapping. The project focused on using commercial image processing and analysis programs where appropriate, to facilitate technology transfer.

3.2. Past SERDP Projects

The Center for Spatial Technologies and Remote Sensing (CSTARS) had no prior experience with SERDP grants.

3.3. Research and Development Proceeding this Project

At the time the study was proposed in 1998, there were no specific examples in the literature for mapping individual species from hyperspectral data. In fact, an influential paper by John Pierce (1994) questioned whether species identification was possible given the phenological changes plants undergo during the growing season, especially when combined with landscape heterogeneity that causes some individuals of a species to be stressed while others are at a fully healthy state. While photogrammetric examples existed in 1998, these were primarily based on visual interpretation and expert knowledge of the interpreter over areas of very limited spatial extent. Similarly, examples using multispectral data were based on classifying bands using clustering algorithms, but observed over a limited range of environmental conditions. These studies were case specific and were generally not applicable in other environmental situations.

Prior to this SERDP study, CSTARS had significant long-term experience in using hyperspectral imagery to map plant community types in a range of ecosystems. CSTARS had also completed a demonstration study for NASA, the U.S. Geologic Survey (USGS), and National Park Service (NPS) on mapping leafy spurge (*Euphorbia esula*) at Theodore Roosevelt National Park in North Dakota (O'Neill et al. 2000) using NASA's Airborne Visible Infrared Imaging Spectrometer (AVIRIS) hyperspectral imagery. This was the first study to map an invasive plant species in a natural landscape, although it built on experience using mixture modeling to map chaparral communities and species in studies conducted with Dar Roberts (1998) of the Santa Monica and Santa Ynez Mountains.

3.4.   Criteria for Detection of Invasive Plant Species

Many plant species are not appropriate for image-based detection because they are indistinguishable from surrounding species or are understory species whose direct detection from remote sensing is almost impossible (Joshi et al. 2004). Successful approaches to detecting invasive species have generally capitalized on the presence of some unique phenological or biochemical properties that are measurable by hyperspectral instruments based on their spectral and structural characteristics. Detection is also relatively easy when the spatial patterns created by the invasive species are large and distinctive monocultures, e.g., leafy spurge, at scales of tens of meters to kilometers.

Spectral characteristics of plants are derived from biochemical absorption features related to chlorophyll and other foliar pigments, water, proteins, starches, waxes, and structural carbohydrate molecules such as lignin and cellulose (Elvidge 1990; Wessman 1992; Fuentes et al. 2001; Penuelas et al. 1997). Plants share a common metabolism, and therefore express a similar suite of biochemical compounds. Identification generally targets differences in the abundance of these biochemicals between species. For example, some biochemical features have little dependence on phenology, such as the strong water-absorption features associated with the succulent leaves of pickleweed (*Salicornia virginica*) (Sanderson et al. 1998) or iceplant (*Carpobrotus edulis*) (Underwood et al., 2003). In contrast, many others depend on capturing a phenological event to identify the unique biochemical attribute, e.g., flower color of the target invader may be distinct, as the case for the yellow bracts of leafy spurge (*Euphorbia esula*) or the unique orange-brown color of Chinese tamarisk (*Tamarix chinensis*) prior to leaf drop. In order to distinguish these features, the timing of image acquisition is of critical importance. It is difficult to time an airborne mission to meet very brief phenological events, as scheduling is often impossible and cloud cover and other factors will limit the times that are suitable for observation. These factors are even more constraining for satellites due to infrequent temporal coverage of high spatial resolution satellites.

3.5.   How Does Remote Sensing Distinguish Invasive Species?

The characteristics of the particular invasive plant species in question determine whether remote sensing techniques are cheaper and more efficient than traditional field-based methods. Successful approaches to detecting invasive species capitalize on identifying unique phenological or biochemical properties that can be measured by hyperspectral imagers based on their spectral and structural characteristics, or their spatial patterns.

3.6.   Biochemical characteristics

Remote sensing images record the reflectance spectra of vegetation (and other land-cover elements such as soils and geologic minerals) within each pixel based on their interactions with electromagnetic radiation in the reflective spectral region. The biochemical and structural properties of these elements determine the observed reflectance. Spectral characteristics of plants are derived from biochemical absorption features related to chlorophyll and other pigments, water, proteins, starches, waxes, and structural carbohydrate molecules such as lignin and cellulose (Elvidge 1990; Fuentes et al. 2001; Penuelas et al. 1997).

Plants share a common metabolism; therefore express a similar suite of biochemical compounds. Identification generally targets differences in the abundance of biochemicals between species. For example as previously described, some biochemical features have strong absorption features, such as those associated with the succulent leaves of pickleweed (*Salicornia virginica*) (Sanderson et al. 1998) or iceplant (*Carpobrotus edulis*) (Underwood et al. 2003). In order to distinguish these features the timing of image acquisition is of critical importance to successful mapping.

3.7.   Ecological Assessment of Native/Invasive Interaction

The ecological research and fieldwork parts of the project were integral in development of the invasive plant species location information and ancillary information on density, plant condition, presence of other species, and other factors like recent disturbance and the type of disturbance (wheel tracks, fire, trampling, erosion, etc.). The study used a combination of field data locating the distribution of invasive plant species and image analysis to develop maps of invasive plant species for the six bases. Typically, as a minimum, the GPS location, species composition, and percent cover were recorded. Sites for data collection were chosen partially from base information about locations of invasive plant species and the dominant vegetation types on the base. Initially, for the SERDP study, more detailed data was collected on plant height, condition, and evidence for disturbance on the plots. One goal of sampling was to acquire some field data at locations across the imagery. Additionally, site information (topography, roads, soils, etc.) was used to identify other areas where invasive plant species were likely to occur. In some cases, due to military activities, some areas had restricted access to various parts of the bases and field crews were unable to fully sample all land cover strata, which in some cases, limited the ability to improve the invasive plant species maps that were produced.

Field spectral measurements were made of the dominant native plant species, invading plant species, and bare soils that were used as reference spectra to create spectral libraries for the image analysis. Later in the study, GPS locations of "patches" were used to identify vegetation and invasive plant species, generally at scales of 5 pixels x 5 pixels in extent (areas of 15-20m x 15-20m in extent), that could be identified in the imagery and used to extract spectra of specific species from the corresponding geolocated pixels. This method turned out to be equally or more effective than field-measured spectra for the 3-4m x 3-4 m (12-16m^2) pixels measured by the airborne AVIRIS data. It was found to be significantly faster, therefore a more cost effective method, than collecting field spectral measurements with a field spectrometer.

The approach to image analysis adopted here was to obtain preliminary GIS information from the base's digital database on topography, road and construction layers, and any vegetation, soil or geology maps available from the base. Additionally, some bases provided information on land use history that helped us identify possible sites of plant species invasion and impact of disturbance. These data sources were used in consultation with base personnel to develop a flight plan for the base that included known locations of the target invasive plant species and areas of the same native vegetation that were not invaded by the invasive plant species. Field sampling locations were chosen to be within the flightlines and include areas of invaded and non-invaded vegetation. This information was used in the GIS to characterize different landscape conditions and used presence and absence information to train the image classifiers and other data to evaluate the accuracy of the map produced in the analysis.

3.8. GIS Model of Invasion

Assuming invasive plant species can be mapped, there are many management questions that a spatial database can help answer. For example, what pre-conditions promote the spread of invasive plant species at the base? How does this relate to military use and activities at the base? Can environmental conditions be identified (e.g., soil type, water flows and drainage, topography, etc.) or human activities (e.g., proximity roads, trails, and construction) that facilitate the spread of invasive plant species? Such information should aid management and identify alternative actions that will reduce spread of invasive species.

At Vandenberg Air Force Base, the information on invasive plant species and native species distributions extracted from the hyperspectral datasets was used to populate a GIS database. This allowed integration with other site characteristics, e.g., soil, topography, and land use, to create a Classification and Regression Tree (CART) model to predict possible locations where the invasive plant species could spread. To a lesser extent this type of CART model was also applied at Fort Benning on the kudzu distribution and on tree decline at the base.

3.9. Technology Transfer

This project created a manual for SERDP to use in personnel training at military bases to assist in performing the hyperspectral image analysis methods that were found to be successful in this study. The training manual and a "User's Guide" for mapping invasive plant species using hyperspectral imagery are provided as an appendix to this report. Once this report is approved, CSTARS will provide each of the six military installations with a copy of the report, training manual, datasets, and User's Guide. It is based on a modification of the tutorial the Environment for Visualization of Images (ENVI), but tailored for the methods found useful in this study. This commercial software program is widely used by the research community for analyzing hyperspectral data. Its pull-down menus are easy to use while the software tools in the program provide advanced capabilities. A workshop was held at UC Davis in 2004 and at least one person from each base received training at UC Davis on the use of the methods that were found most efficient for detecting invasive plant species. The project has provided an updated training manual to SERDP, based on feedback received after the SERDP workshop and a second workshop given to graduate students and others at Davis on mapping invasive species.

4. Methods and Materials

4.1. Methods for Overall Project

Hyperspectral imagery was used to map the following invasive plant species at the following six bases (Table 1).

Table 1: Study sites, Target Invasive Plant Species and Associated Habitat.

Study Site	Target Invasive Plant Species	Habitat
Vandenberg Air Force Base	Iceplant - *Carpobrotus* spp. (*C. edulis*, *C. chilensis*, hybrids)	Coastal Dunes
Vandenberg Air Force Base	Jubata Grass	Chaparral
Camp Pendleton Marine Corp Base	Fennel - *Foeniculum vulgare* Giant reed - *Arundo donax*	riparian
Fort Benning	Lovegrass - *Eragrostis* spp. Kudzu – *Pueraria lobata*	Deciduos Forest
Yuma Proving Ground	Mediterranean grass - *Schismus* spp. Asian mustard *Brassica tournefortii* Tamarisk - *Tamarix* spp. (*T. chinesis, T. aphylla*)	Sonoran Desert
Yakima Training Center	Cheatgrass – *Bromus tectorum* Russian knapweed plant species - *Acroptilon repens*	Great Basin Sagebrush
Aberdeen Proving Ground	Common reed – *Phragmites australis* Hydrilla – *Hydrilla verticillata*	Riparian

4.2. Brief Description of Sites and Invasive Plant Species Studied

4.2.1. Vandenberg Air Force Base, California

This base is located in the Pacific coastal margin physiographic province with a strong Mediterranean climate, in a climate transition zone between northern and southern California. It occurs on diverse landforms including dunes, hilly terrain and wetlands that are home to several endangered plant and animal species. This study examined invasive plant species in two dominant southern California forms of low-stature shrub ecosystem types, the Coastal Sage Scrub, a diverse and globally rare habitat type occurring in coastal terraces and foothills below 1,000 m occupying the dunes and coastal margin at Vandenberg and the Burton Mesa Mixed Chaparral. The Coastal Sage Scrub type is termed soft chaparral due to the predominance of soft, drought-deciduous leaves in contrast to the hard, waxy-cuticle leaves characteristic of the sclerophyllous plants of California's chaparral communities. Characteristic plants include California sagebrush (*Artemisia californica*), black sage (*Salvia mellifera*), white sage (*Salvia apiana*), California buckwheat (*Eriogonum fasciculatum*), coast brittle-bush (*Encelia californica*), golden yarrow (*Eriophyllum confertifolium*), with the larger shrubs toyon (*Heteromeles arbutifolia*) and lemonade berry (*Rhus integrifolia*), along with other shrubs and herbaceous plants, grasses, and in some places, cacti and succulents.

The California Gnatcatcher *(Polioptila californica)*, is a critically endangered bird species endemic to this type and 13 plant species are recognized as threatened or endangered. The Burton Mesa Mixed Chaparral is a unique form of maritime chaparral restricted to the aeolian sands of the Orcutt Formation, a few miles inland from the Pacific Ocean. This sclerophyllous chaparral is characterized by numerous endemic species, many found only in this community. The dominants include manzanitas (*Arctostaphylos rudis* and *A. purissima*), varieties of Lompoc and Santa Barbara ceanothus (Ceanothus cuneatus var. fascicularis and *C. impressus* var. *impressus*), local varieties of flowering annuals and many other taxa. Both vegetation types represent unique and endangered ecosystem types on the California coast.

1. Iceplant (*Carpobrotus edulis*) – an aggressive invader into endangered Coastal Sage Scrub that alters soil salinity and moisture, allowing it to form extensive monotypic stands. Has succulent prostrate herbaceous species with bright red/yellow flowers; succulent leaves turn red in fall, increasing detectability at that time. Causes changes in biochemical cycling and soil surface salt accumulation; restricts regrowth of native species.

2. Jubata grass (*Cortaderia jubata*) – an aggressive invader into coastal mixed chaparral and forms monotypic stands, taller than native shrub species. Has a tall bunchgrass habit with large white terminal inflorescences and accumulation of dry leaves in canopy. Jubata grass provides a comparison with giant reed and common reed in terms of structure, but grows in open sparse shrublands while the two bamboo-like grasses are riparian and fresh to brackish marsh species.

4.2.2.   Camp Pendleton Marine Base, California

This base is located in the southern California Pacific coastal margin physiographic province having a strong Mediterranean climate. Camp Pendleton extends over approximately 53,872 ha, with coastal dunes and lowlands, and wetlands in the southern-most part of the Base. The steep west-facing hillslopes extend from the California coastal range along its eastern border to its ocean border. Along the coastal margin, Camp Pendleton retains one of the largest remnants of relatively intact Coastal Sage Scrub on the southern California coast. Although much of the hillslopes of the base are composed of extensively disturbed and fragmented patches of semi-arid grasslands, mixed chaparral, and along the Santa Margarita River and the other smaller creeks that flow through the base to the Pacific Ocean, riparian woodlands. Riparian woodlands support several species of willow, cottonwoods, sycamore, coast live oak, ash, white alder, and a diverse flora of herbaceous plants, shrubs, and vines, including several listed and endangered species.

1. Giant reed (*Arundo donax*) –a very aggressive riparian grass species that forms dense, tall cane (up to 7 m height) or bamboo-like monospecific stands. Canopy is distinctive due to its height and density. It changes hydrology and riparian margin ecology; due to its extensive water use during the dry season, producing abundant biomass that promotes more frequent wildfire in riparian zone during dry season.

2. Fennel (*Foeniculum vulgare*) –an aggressive invader into grasslands of the Coastal Sage Scrub and mixed chaparral semiarid shrublands of southern California. This herbaceous species is difficult to measure from remote sensing due to the diffuse and sparse canopy and low leaf cover. This species presents a significant detectability challenge for the analysis.

4.2.3. Fort Benning, Georgia

Fort Benning occupies 73,653 ha on the Chattahoochee River, along the Alabama-Georgia border, on the linear transition zone between the higher Piedmont physiographic province to the north and west, and the lower Coastal Plain physiographic province to the south and east. The Southern Mixed Forest is composed of mixed pine on sandy soils interspersed with oak-hickory or beech-magolia hardwood forests that include several oak species and sweetgum. Ft. Benning is a national heritage site; species diversity is high and there are 66 known State and Federal Threatened, Endangered, and Special Concern plant and animal species on the base. The detection problem at this site is to identify a single broadleaf invasive plant species within the variable composition of a mixed forest class.

1. Kudzu (*Pueraria lobata*) –aggressive invasive liana using other species for support invading open disturbed sites, in forests, riparian zones, roadsides, and in abandoned fields or logged forest. Kudzu is common in riparian zones in the eastern deciduous forest and southeastern mixed pine forest biomes. Leafy crown grows over native vegetation, cutting off light and eventually killing the overgrown species. While distinctive by sight from ground, it is likely to be difficult to separate its foliage from the native vegetation it grows over in remote sensing data. The timing of its annual growth is largely in phenological synchrony with background vegetation, also contributing to difficulty in identification.

2. Weeping lovegrass (*Eragrostis curvula*) – this species has potential to be invasive and is under study at Ft. Benning to determine its risk. The base has planted this grass species extensively for erosion control, a significant environmental management problem. As such, the base has identified this species as a top priority. This species is primarily planted along roads or on open land patches but could expand into the understory of the surrounding forest. Because the base knows the locations where it has been planted for erosion control (even if they cannot provide all locations), these data can be used to train a detection algorithm. Some sites may be planted in the understory of native vegetation and be undetectable or too sparse to be detected.

4.2.4. Yuma Proving Ground, Arizona

This base, 336,700 ha established in 1943 for weapons testing, is located in southwestern Arizona in the hot Sonoran Desert physiographic province within the North American monsoon climate zone. The bi-seasonal rainfall pattern, with mild winter storms and intense summer thunderstorms, combined with a low total amount of annual rainfall define the characteristics of the Sonoran Desert. The Yuma desert is one of the most arid parts of the U.S., with low-stature low-density xeric shrubs, primarily composed of widely spaced creosote bush (*Larrea divaricata*) and white bursage (*Ambrosia dumosa*), which are among the most drought tolerant species known, and a limited diversity of other Sonoran shrubs. These species typically respond to winter precipitation and only grow abundantly in wet years. Along the washes and riparian zones canopy cover is higher and palo verde (*Cercidium floridum, C. microphyllum*) and ironwood (*Olneya tesota*), saguaros (*Carnegiea gigantia*), ocotillo (*Fouquieria splendens*), elephant trees (*Bursera* spp.), and limberbush (*Jatropha* spp.) are common but which are often leafless unless water is available. Two dominant plant life forms distinguish this desert from other North American deserts, the tree legumes like palo verde and columnar saguaro cacti. The combination of growth forms, low plant density and low leaf cover of perennial species make

this site challenging to detect invasive species with hyperspectral imagery. Cryptophytes and annuals grow for brief periods following the infrequent and unpredictable rains. Annual species generally comprise more than half the plant diversity (up to 90% in the driest sites). The spatial and temporal variability of the growth of annuals, particularly in response to summer rainfall, contribute additional complexity to the analysis.

1. Salt cedar (*Tamarix* species) - aggressive phreatophyic riparian zone shrub invader, changing hydrology, fire cycle, and nutrient cycling of the habitat. Large shrub/sub-tree form growing with willows and riparian trees along water courses and in other semiarid sites where water is available. *Tamarix* out-competes native vegetation for water resources allowing rapid growth and accumulation of woody biomass, which enhances fire frequency. Co-occurrence of palo verde and tamarisk in the riparian zone of desert washes allows evaluation of whether, given their similar growth forms, they can be distinguished in hyperspectral imagery.

2. There was an intention to attempt to map the invasive Mediterranean grass, *Schismus* spp. and Asian mustard, *Brassica tournefortii,* however due to the severe multi-year extended drought in the southwest during the timing of the overflights for this project, these species were dormant or died after very little growth. Thus, it was impossible to test the potential for hyperspectral imaging to detect these species.

4.2.5. Yakima Training Center, Washington

The 100,000 ha Yakima Training Center is located in the upper part of the Yakima Valley, trending east-west and representative cold desert Great Basin physiographic province. Topography is complex with numerous minor valleys and ridges giving a variation in altitude of up to 0.6 km. This variation in local relief creates significant variation in microclimate over short distances. The valley has relatively hot, dry summers and cool, mild winters with light snowfall. Several steppe shrub communities occur within zones determined by microclimate and topographic conditions. Vegetation is dominated by a few species of low, small-leafed shrubs; with few trees, succulents or annuals. The dominant community is the big sagebrush (*Artemisia tridentata*) - bluebunch wheatgrass (*Agropyrons spicatum*) association, which often extends over large areas with shrub density typically at 25% cover or less. The detection problem is to distinguish an invasive grass or herbaceous species at the sub-pixel scale within the shrub community.

1. Cheatgrass (*Bromus tectorum*) – grass forming widespread patchy or monoculture infestations replacing Great Basin sagebrush scrub and bunchgrass scrub. It can grow in the shrub understory, creating a continuous fuel distribution for wildfire spread. Frequent wildfires reduce shrub replacement and promote long-term conversion of shrubland into grassland. The difficulty of the remote sensing observation is to distinguish cheatgrass from native grasses and shrubs.

2. Russian knapweed (*Acroptilon repens*) – aggressive herbaceous species that forms a monoculture in Great Basin sagebrush scrub and stays green later than surrounding vegetation. Its ecology is also closely tied to the fire cycle.

4.2.6. Aberdeen Proving Ground

Aberdeen Proving Ground, established in 1917, is an ordinance testing base of 12,141 ha and adjacent open water, located on the northwestern shore of the Chesapeake Bay. It has four distinct climate seasons, with maximum rainfall in summer and minimum in autumn. It is drained by eight rivers and streams, with tidal estuaries forming at their mouths and wetlands characterized by fluctuating water levels and variable salinity. Brackish wetlands have richer species diversity than salt marshes, and are largely composed of grasses, sedges, rushes, aster, and chenopod species, with distributions largely controlled by topography and tidal flooding. Common brackish species include *Iva frutescens, Typha angustifolia, Scirpus olnei, S. robustus, Hibiscus palustris, Eryngium virginianum.*

1. Common reed (*Phragmites australis*) - large cane or bamboo-like grass, similar in growth form but smaller than *Arundo donax*. It grows in riparian and disturbed habitats with altered hydrology and topography and eutrophied fresh and brackish water. The structure of common reed allows comparison to spectral detection of *Arundo donax* at Camp Pendleton.

2. Hydrilla (*Hydrilla verticillata*) – aggressive submerged aquatic herbaceous species growing in estuaries and fresh to brackish water habitats. The difficulty is to determine presence when flooded under variable tide levels and in turbid water.

4.3.   Remote Sensing Methods Used in this Study

This section describes the methods that were found to be most consistently useful in mapping invasive plant species at the military installations. A summary of the methods that were best in detecting each of the invasive plant species for the site studies are shown in Table 2. Past experience in other studies supports the selection of these methods as the most commonly useful techniques. The descriptions of these methods, provided below, have been adapted from the Environment for Visualization of Images (ENVI) tutorial manual (Research Systems, 2002).

4.3.1.   Spectral Angle Mapper

The Spectral Angle Mapper (SAM™, ENVI, Remote Systems, Inc.) is an automated method for comparing spectra of image pixels to the spectra of reference endmembers (also termed "training classes") (CSES, 1992; Kruse *et al*., 1993). SAM treats each spectrum as a vector and determines the similarity between the unknown pixel spectrum and the reference spectrum by calculating the "spectral angle" between them. A simple example is shown in Figure 1, which uses a reference spectrum (material A) and an unknown spectrum (material B) from a data set of *n* spectral bands. The reflectance of each material can be plotted in a 2-D scatter plot for each combination of bands in the data set. A vector is then extended from the origin to the point of I-J reflectance for each material. The 2-band spectral angle for the reference and unknown pixels in the I-J data space is the angle between the two vectors.

The color of a material is defined by the direction of its unit vector in *n*-dimensional spectral space. The length of the vector relates only to how fully the pixel is illuminated, not its absorption features. Because SAM uses only the direction of the spectrum, and not its length, all

levels of illumination are treated equally. Poorly illuminated pixels fall closer to the origin, but still lie on the same vector as brightly illuminated pixels. Notice that the angle between the vectors is the same regardless of the length.

Table 2: Site and species characteristics, and preferred mapping methods

Base	State	Invasive plant species	Environment	Species traits	Best mapping method
Vandenberg	CA	Iceplant	Mediterranean, semi-arid chaparral	Succulent	ML on MNF
Vandenberg	CA	Jubata Grass	Mediterranean, semi-arid chaparral	Grass	ML on MNF
Camp Pendleton	CA	Fennel	Mediterranean, semi-arid chaparral	Sparse canopy herbaceous	MTMF with indexes
Camp Pendleton	CA	Giant cane	Riparian	Robust reed	ML on MNF
Fort Benning	GA	Kudzu	Temperate forest	Vine	SAM on MNF
Fort Benning	GA	Love Grass	Temperate meadows	Grass	SAM on MNF
Yuma	AZ	Tamarisk – irrigated	Irrigated desert	Dense canopy shrub	Any
Yuma	AZ	Tamarisk – wash	Desert	Sparse canopy shrub	MTMF
Yakima	WA	Cheatgrass	Semi-arid grassland	Green & senescent grass	MTMF
Yakima	WA	Russian Knapweed	Semi-arid grassland	Sparse canopy herbaceous	MTMF
Aberdeen	MD	Common cane – Carroll	Salt marsh	Dense reed	SAM on indexes
Aberdeen	MD	Common cane - Spesutie	Salt marsh	Mixed reed	MTMF
Aberdeen	MD	Hydrilla	Ocean shallows	Submerged aquatic vegetation	-
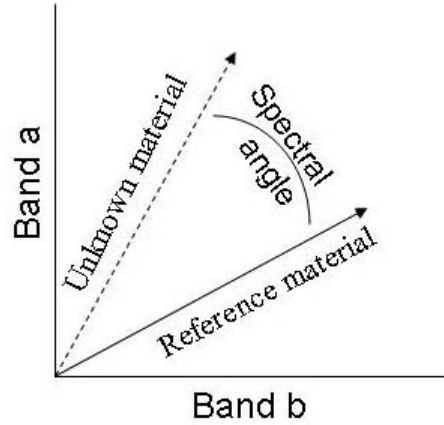
Figure 1: Two-dimensional example of the Spectral Angle Mapper

The SAM algorithm generalizes this geometric interpretation to *n*-dimensional space. SAM determines the similarity of an unknown spectrum *t* to a reference spectrum *r* by applying Equation 1 (CSES, 1992):

$$\alpha = \cos^{-1}\left(\frac{\vec{t} \cdot \vec{r}}{\|\vec{t}\| \cdot \|\vec{r}\|}\right) \qquad [1]$$

which can also be written as:

$$\alpha = \cos^{-1}\left(\frac{\sum_{i=1}^{n} t_i r_i}{\left(\sum_{i=1}^{n} t_i^2\right)^{\frac{1}{2}}\left(\sum_{i=1}^{n} r_i^2\right)^{\frac{1}{2}}}\right) \qquad [2]$$

where *n* equals the number of bands in the image.

For each reference spectrum (also termed "endmember", "training class", or "pure type") used in the analysis of a hyperspectral image, the spectral angle α is determined for each pixel spectrum in the image. This value, in radians, is assigned to the corresponding pixel in the output SAM image, with one output image for each endmember analyzed. The derived spectral angle maps form a new data cube with the number of bands equal to the number of reference spectra used in the mapping. The user specifies a threshold angle which identifies the maximum allowable angle difference between the reference and the unknown pixel. The pixel will be assigned to the reference class that it is most similar to, i.e., the one with the smallest spectral angle. If α exceeds the threshold angle, the unknown will not be placed in the reference class and if it does not fit any reference class it is identified as "undefined".

The SAM algorithm was very useful for species and vegetation type mapping and it has been included it as part of the "standard" image processing package for the tutorial and for project data analysis. The method works best if the vegetation classes are well separated spectrally, but less well with poorly separated classes.

4.3.2. Spectral Indexes

A large number of spectral indexes have been identified for use with both multispectral and hyperspectral datasets. It is an important data reduction technique that can reduce the number of bands used in an analysis but retain the most important spectral information, by calculating spectral indexes relevant to plant properties and functions.

Spectral indexes, in their simplest form, are ratios or linear combinations of two or more spectral bands (Jackson & Huete 1991), typically one band at a wavelength affected by the absorbing material of interest and one at a reference wavelength where the material does not absorb. As shown in Figure 2, leaf reflectance spectra are dominated by the absorbance of chlorophyll and water. Not surprisingly, spectral indexes are frequently derived to estimate these parameters. The most common are vegetation indexes, especially the normalized difference vegetation index (NDVI), which has been used in numerous published studies (> 2000 found searching ISI Web of Science). These indexes are sensitive to green plant cover and provide an estimate of the "greenness" of an area. As such, they have shown to be correlates of percent cover of vegetation, leaf area index (LAI), and plant biomass.
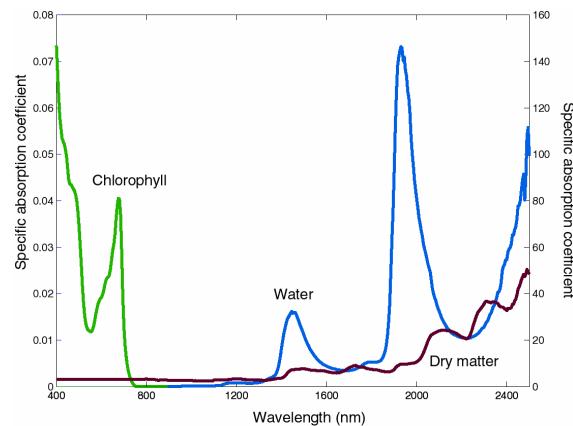
Figure 2: Specific absorption coefficients for absorption in leaves by chlorophyll a+b (cm2 μg^{-1}) on the left axis, of water (cm^{-1}) and dry matter (cm^2 g$^-$) on the right axis (from Ustin & Jacquemoud 2002).

Hyperspectral instruments allow the elaboration of spectral indexes beyond simple greenness indicators. Since AVIRIS samples 224 bands between 400 and 2500 nm, it allows the detection of narrow absorption features. High spectral resolution instruments have used these narrow features to elucidate foliar chemistry such as lignin, starch, protein and cellulose concentrations (Card et al. 1988, Fourty & Baret 1998, Curran 1989). These techniques have proven so successful in laboratory settings that they can replace traditional wet chemistry extraction methods (Curran 1989). The detectability of foliar chemistry suggests a potential for developing spectral indexes sensitive to subtle variations in leaf physiology. Numerous physiological indexes now exist for detecting a range of leaf properties.

Physiological index based classifications allow for explicit interpretations of physical properties and, therefore, can be preferable to typical statistical classifications for which the basis of the differences in an image data set are unknown. Statistical classifications integrate all information contained within class spectra and capitalize on the fact that classes are different, but do not

13

provide any information regarding the basis of that differentiation. Also, principal components based methods characterize differences based on the variance in the image dataset and, as such, are dataset dependent and cannot be extrapolated to other datasets. Furthermore, these exploited differences may be habitat variables that only indirectly relate to vegetation cover or the explicit features of the species. Physiological indexes do not have this limitation because they estimate physical parameters to differentiate classes and, as such, have potential for comparison to other datasets. For example, Underwood et al. (2003) successfully mapped dense monotypic infestations of invasive iceplant at Vandenberg AFB using a simple water band index (Ustin et al., 1998), indicating that physiological differences (in this case succulence contributing to foliar water content) confer spectral uniqueness.

The utility of physiological indexes was investigated for mapping invasive species at all six bases. Of particular interest has been the generality of indexes across sites and conditions. Do the same collection of indexes function well across ecological conditions, or are sets of indexes site- or species-specific? Of the possible indexes, 19 physiological indexes were used of three basic types: pigment indexes, foliar water indexes, and foliar chemistry indexes (Table 3 and Table 4). The general conclusion obtained is that within each of the three general classes of indexes (pigments, water, and other foliar chemistry), due to correlations, indexes generally showed the same pattern, although the strength of the correlation varied between these indexes. When several indexes were combined in a classification, there was sufficient variability to produce a classification that did not mirror any one index.

Pigment indexes focus on the absorptions of photosynthetic pigments. Several specifically highlight chlorophyll, the predominant photosynthetic pigment of green vegetation. In addition to chlorophyll, however, plants possess other accessory pigments, e.g., several carotenoid pigments like violaxanthin and zeathanthin. The basis for using these indexes in a classification is that the relative values measured using these indexes is often consistent for a species or a vegetation type. However, these patterns are not always conclusive and relative accessory pigment concentrations may vary between species or between individuals of the same species experiencing different stress levels (Zarco-Tejada 1998), therefore differences must be interpreted carefully. See Table 3 for a listing of pigment indexes studied including descriptions of their physiological sensitivities.

Water absorbs strongly throughout the solar infrared wavelengths. Foliar water content thus heavily influences reflectance throughout these regions (Carter 1991). In particular, vegetation spectra exhibit strong water absorption features at 0.97, 1.24, 1.4, and 1.9 µm (note units that 970 nm = 0.97 µm, or a factor of 1000 difference; optical remote sensing typically uses both wavelength units interchangeably). The depth of these absorption features corresponds to foliar water content, although not linearly. It is possible, using theoretical curves for water absorption, to relate the depth and breadth of the absorption feature to concentration of water. Invasive species with strategies of water uptake (e.g, tamarisk) or storage (e.g, iceplant) that differ from native vegetation are expected to display water index values that reflect these strategies, and thus be readily detectable. An example of a water index and difference between a native species and invasive plant species is shown in Figure 3.

Table 3: Pigment based physiological indexes used in vegetation mapping.

Index	formula	Details	citation
Pigment indexes			
SR, Simple Ratio	$\dfrac{R_{NIR}}{R_R}$	Index of green vegetation cover. Various wavelengths used, depending on sensor.(eg: NIR=845nm, R=665nm)	Tucker (1979)
NDVI, Normalized Difference Vegetation Index	$\dfrac{R_{NIR} - R_R}{R_{NIR} + R_R}$	Index of green vegetation cover. Various wavelengths used, depending on sensor.(eg: NIR=845nm, R=665nm)	Tucker (1979)
mNDVI, modified NDVI	$\dfrac{R_{750} - R_{705}}{R_{750} + R_{705}}$	leaf chlorophyll content	Fuentes et al. (2001)
Summed green reflectance	$\displaystyle\sum_{n=500}^{599} R_n$	Index of green vegetation cover.	Fuentes et al. (2001)
PRI, Photochemical Reflectance Index	$\dfrac{R_{531} - R_{570}}{R_{531} + R_{570}}$	Xanthophyll response to light ~ photosynthetic efficiency. Also sensitive to carotenoid/chlorophyll ratio	Rahman et al. (2001)
Red/Green ratio	$\dfrac{R_{600-699}}{R_{500-599}}$	anthocyanins/chlorophyll	Fuentes et al. (2001)
NPCI, Normalized Pigments Chlorophyll Ratio Index	$\dfrac{R_{680} - R_{430}}{R_{680} + R_{430}}$	total pigments/chlorophyll	Peñuelas et al. (1995)
SRPI, Simple Ratio Pigment Index	$\dfrac{R_{430}}{R_{680}}$	carotenoid/chlorophyll *a* content	Zarco-Tejada (1998)
NPQI, Normalized Phaeophytinization Index	$\dfrac{R_{415} - R_{435}}{R_{415} + R_{435}}$	chlorophyll degradation, detects stress at early states	Zarco-Tejada (1998)
SIPI, Structure Intensive Pigment Index	$\dfrac{R_{800} - R_{445}}{R_{800} - R_{680}}$	carotenoid/chlorophyll *a* concentrations	Zarco-Tejada (1998)
PI1, Pigment Index 1	$\dfrac{R_{695}}{R_{420}}$	plant stress status	Zarco-Tejada (1998)
PI2, Pigment Index 2	$\dfrac{R_{695}}{R_{760}}$	plant stress status	Zarco-Tejada (1998)
PI3, Pigment Index 3	$\dfrac{R_{440}}{R_{690}}$	vegetation health, based on chlorophyll fluorescence ratios	Lichtenthaler et al. (1996)
PI4, Pigment Index 4	$\dfrac{R_{440}}{R_{740}}$	vegetation health, based on chlorophyll fluorescence ratios	Lichtenthaler et al. (1996)

Table 4: Water and other biochemical physiological indexes used in vegetation mapping.

Index	formula	Details	citation
Water indexes			
NDWI, Normalized Difference Water Index	$\dfrac{R_{860} - R_{1240}}{R_{860} + R_{1240}}$	leaf water content	Gao (1996)
WBI, Water Band Index	$\dfrac{R_{900}}{R_{970}}$	leaf water content	Peñuelas et al. (1997)
Foliar chemistry indexes			
NDNI, Normalized Difference Nitrogen Index	$\dfrac{\log\left(R_{1680}/R_{1510}\right)}{\log\left(1/R_{1680}R_{1510}\right)}$	foliar nitrogen concentration	Serrano et al. (2002)
NDLI, Normalized Difference Lignin Index	$\dfrac{\log\left(R_{1680}/R_{1754}\right)}{\log\left(1/R_{1680}R_{1754}\right)}$	foliar lignin concentration	Serrano et al. (2002)
CAI, Cellulose Absorption Index	$0.5 * (R_{2020}+R_{2220})- R_{2100}$	based upon cellulose & lignin absorption features, used to discriminate plant litter from soils	Nagler et al. (2000)

Figure 3: The bands used by the water band index (WBI) for sample vegetation at VAFB (left) and Yuma Proving Ground (right).

Finally, several indexes have been proposed to estimate leaf chemistry. As with pigment and water indexes, it is probable that different species display distinct proportions of foliar chemicals. For example, the invasive vine kudzu is leguminous, forming a symbiotic association with nitrogen fixing root nodules. As a result, it may have higher foliar nitrogen content than native species, and consequently higher normalized difference nitrogen index (NDNI) values. Note that the cellulose absorbance index (CAI; Nagler et al. 2000), in contrast to the other indexes studied, estimates the cellulose content of senescent plant matter, emphasizing the

distinctions between soil and plant litter. This index has been used in riparian zones in semi-arid environments and may be most useful where healthy green vegetation is (seasonally) sparse or to detect fall senescence in mixed evergreen and deciduous ecosystems.

All of the parameters estimated by the physiological indexes described here are in leaves, and as a result, all tend to be correlated through leaf area. Obviously the several indexes estimating chlorophyll content will be related. Additionally, there are correlations between various physiological parameters within plants, leading to further correlations between physiological indexes. For example, the water indexes and NDNI are generally highly correlated with NDVI. This is because the more water and nitrogen a plant has access to, resulting in higher leaf water and nitrogen content, the healthier it is and the more it is able to invest in photosynthetic machinery, resulting in higher NDVI. Of interest are the slight differences in spectral leaf traits across species and environments.

The physiological index composite images are created by deriving a variance/covariance matrix, eliminating the most highly correlated bands, and then "stacking" the single band indexes into a final multi-band image. These index images can be used to encapsulate the information contained within vegetation spectra in meaningful and interpretable ways. Physiological indexes provide several benefits, including relatively straightforward interpretation of spatial patterns and differences between species. Index methods are portable across species and sites because of their interpretability and the fact that the index value is not dependent on overall variance as are the principal component based methods.

4.3.3. Continuum Removal

Continuum removal is a procedure that reduces the data volume by focusing the analysis on bands within specific absorption features. It is a process that compares absorption features of pixel spectra to a background reference. The "continuum" is a representation of what a reflectance spectrum would look like if no narrowband specific absorptions were present. It is created by fitting a convex hull to the pixel reflectance spectrum using straight line segments to connect local spectra maxima that span the bands of an absorption feature, as shown in Figure 4.
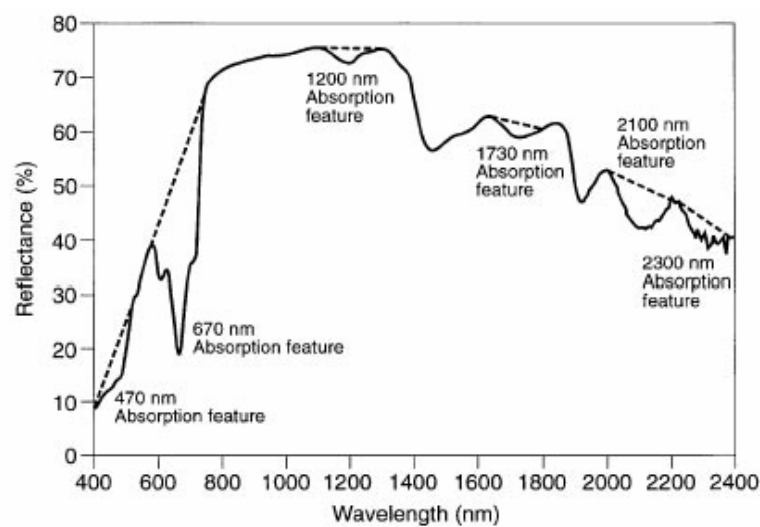
Figure 4: Example spectrum showing common absorption features and connecting line segments for continuum removal analysis

The continuum line segments provide a common baseline from which to compare individual absorption features. This fitting to the background reflectance can be described as fitting "shrink wrap" to the spectrum.

The continuum reflectance, R_c, is removed by dividing it by the actual reflectance, R, for each pixel in the image (Figure 5).

$$r' = \frac{R}{R_c}$$

[3]

The resulting image spectrum, /r'/, is equal to 1.0 where the continuum and the spectrum have equal reflectance, and less than 1.0 where absorption features occur, as shown in Figure 5. This figure shows typical plant and soil absorption features as noted in the lower reflectance spectrum. The upper spectrum shows the same features rescaled as a fractional % of the continuum (background) reflectance. This normalization and rescaling relative to the background allows quantitative comparison of data.
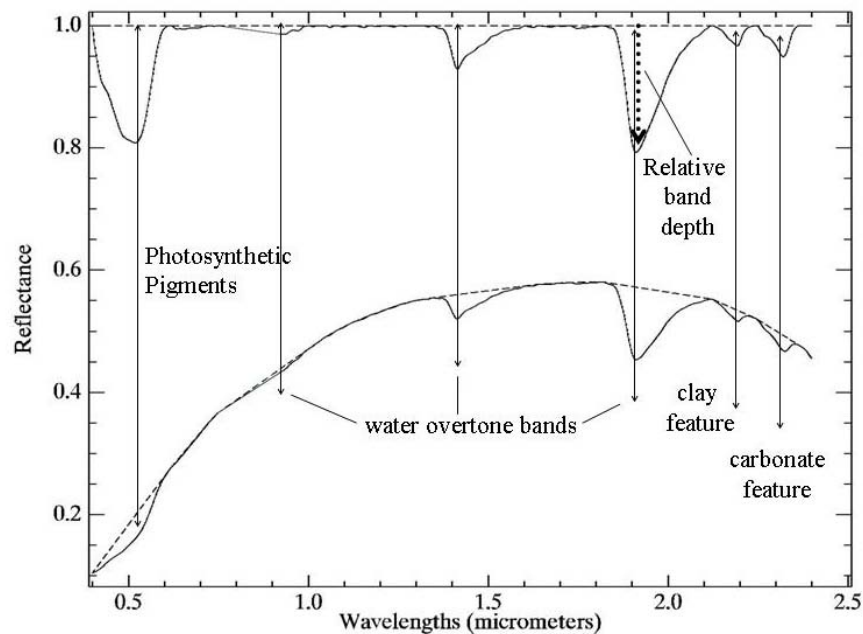
Figure 5: Example of a continuum removal calculation for a mixed pixel spectrum showing typical plant and soil absorption features. The lower figure shows the reflectance spectrum of a pixel while the upper figure shows the same data with the continuum removed, leaving the individual absorption features as deviations from the background reflectance.

The continuum removal technique works best when the target invasive plant species have absorption features significantly different (by wavelength or depth) than the background pixels. This technique may be particularly appropriate in differentiating between the following classes.

- Succulent vs. non-succulent vegetation
- Green vegetation vs. dead vegetation
- Vegetation vs. soil

The absorption feature is selected after a review of the image spectra and the hyperspectral data is subset to the region containing the absorption features of interest.

4.3.4.   Minimum Noise Fraction

The minimum noise fraction (MNF™, ENVI, Remote Systems, Inc.) transformation is a two step principal component analysis, that is used to whiten the noise (i.e., reduce wavelength specific noise) and to reduce the dimensionality of the hyperspectral data (compress the data into fewer "statistically significant" bands). The MNF transform as modified from Green et al. (1988) and implemented in ENVI, is essentially two cascaded Principal Components Analysis (PCA) transformations.

The PCA is a transformation method for distilling the data by reducing the dimensionality of the data set (i.e., a 224 spectral band data set has 224 spectral dimensions). The image data as acquired from the sensor is a data cloud in multi-dimensional space with each band generating an axis of dimension. When the data cloud is viewed in two or three dimensions, the shape of the cloud depends on the bands chosen to supply the axis of the visual plot. Some bands are highly correlated and if they are chosen as axes there is very little variation in the data plot. The PCA is a linear transformation that re-orients the axes to capture the spectral information in un-correlated (orthogonal) axes. The resultant transformed axes are linear combinations of the previous (spectral band based) axes, where each original band contributes to the new transformed band, proportional to the variance it explains.

The PCA does the axis-transformation by performing an eigenvector decomposition of the sample covariance of the data. The first axis is placed on a trajectory to capture the widest variation in the data cloud. The second axis is orthogonal to the first and captures the second widest variation and so forth. Figure 6 is a schematic demonstration of the first two axes.
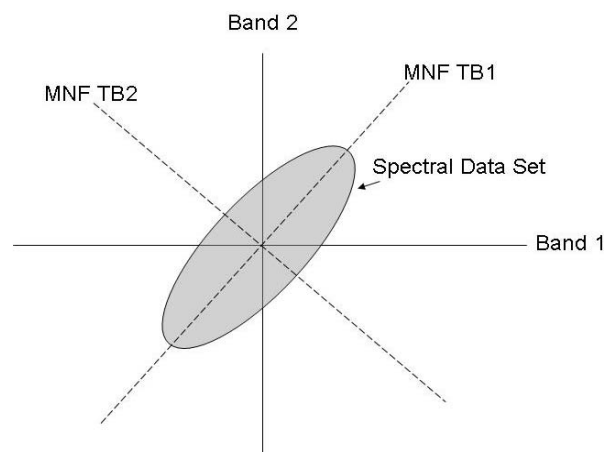
Figure 6: Representation of the first two eigenvectors from a Minimum Noise Fraction decomposition from the distribution of a hypothetical spectral data set. The figure shows the original spectral bands as band 1 and band 2 and the transformed bands as MNF TB1 a d MNF TB2.

The disadvantage in using the PCA reduction in image analysis is that the eigenvectors are based on data variance and not on the signal-to-noise ratio. Since noise is detector sensitive and the solar signal is wavelength dependent (and band based) these sources of noise can contribute to large data variances, making this method unable to reliably separate signal from noise. The MNF resolves this problem by using a two-stage PCA. The first transformation, is based on an estimated noise covariance matrix that de-correlates and rescales the noise (whitens or spreads the noise across all bands) in the data. The second step is a standard PCA transformation of the noise-whitened data. The inherent dimensionality of the data is determined by examination of the final eigenvalues and the associated images.

MNF data space can be divided into two parts: coherent eigenimages and noise-dominated images. By using only the coherent orthogonal images, and eliminating transformed bands that explain little of the variance, the noise is separated from the useful data. The MNF thus improves spectral processing results and substantially reduces the volume of the image data set.
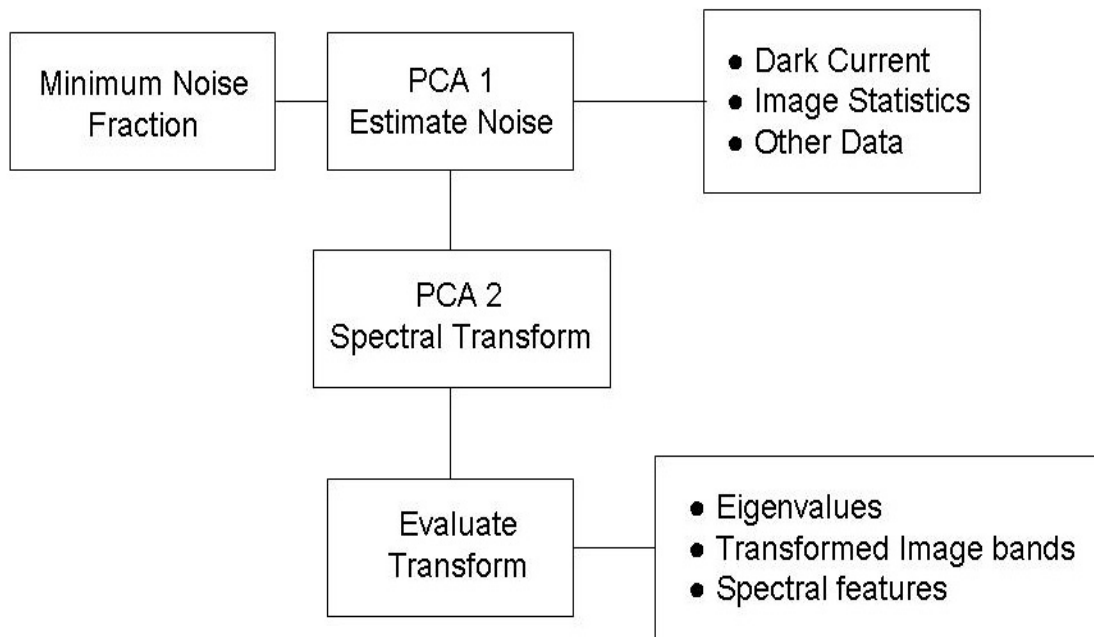
Figure 7: MNF Procedures in the ENVI program (Research Systems, 2002).

Figure 7 summarizes the MNF procedure in ENVI. The noise estimate can come from one of three sources; from the dark current image acquired with the AVIRIS data, from noise statistics calculated from the image, or from statistics saved from a previous transform. Both the eigenvalues and the MNF images (eigenimages) are used to evaluate the dimensionality of the data. Eigenvalues for transformed bands that contain information are typically an order of magnitude larger than those that contain only noise. The corresponding images are spatially coherent when viewed by the user, with little speckle, while the noise images will not contain any discernable spatial information.

4.3.5.   Analyzing at the Sub-Pixel Level

The analyses discussed previously are used to assign pixels to classes, with one class per pixel. However, pixels are rarely pure (i.e., one material only) and each pixel contains a multitude of objects. The previous classification techniques address this by defining mixed classes, where the spectral center of the cluster is the linear combination of the spectra of the individual objects. The classification of a pixel is dependent on the dominant spectral feature of the pixel and the definition of the classes. The target invasive plant species is classified both as an individual class where the pixel is dominated by the invasive species, and as part of a dominant vegetation class (e.g., chaparral-iceplant class at VAFB).

This technique works well for invasive plant species that grow in thick clusters and produce fairly strong signals within the pixel. However, if the invasive plant species of concern does not dominate a feature in the pixels used (from 9 to12 meter square pixels in this study), it may not be identified as a class in the classification. The next higher level of mapping techniques assigns multiple classes to pixels and estimates the portion that each class contributes to the overall spectrum of the pixel.

4.3.6.   Mixing Models

Spectral mixing occurs when materials with different spectral properties are represented within a single image pixel. The simplest model of a mixed spectrum is a linear one, in which the spectrum is a linear combination of the pure spectra of the materials located in the pixel area, weighted by their fractional abundance. The linear spectral unmixing model assumes no interaction between materials in a pixel. If each photon sees only one material, these signals are additive (a linear process). This method approximates the abundance of materials in mixtures, but this assumption does work well in some circumstances (Boardman and Kruse, 1994). There are many published papers that have successfully used this technique (more than 3000 papers cited on ISI Web of Science for a range of image and non-imaged based "spectral mixtures"). Spectral unmixing is one of the most widely used methods in hyperspectral image analysis since the fractional abundance is based on a least-squares fit across all bands.

Mixed pixels can be visualized as points in *n*-dimensional scatter-plot space (spectral space), where *n* is the number of bands. If only two endmembers mix in two dimensions, the mixed pixels will fall along a line (Figure 8A). The pure endmembers are the two ends of the mixing line. If three endmembers mix, then the mixed pixels fall inside a triangle (Figure 8B) and the endmembers are the vertices. Higher dimensions of mixing are represented by higher dimension geometric figures (e.g.: four endmembers mix within a tetrahedron, etc.).

Mixtures of the endmembers fill in the space between the endmembers. All mixed spectra are interior to the convex hull that defines the location of pure endmembers. Mixtures are inside the simplex formed by the endmember vertices because all pixel abundances are positive and sum to unity in a "real" mixture. Mathematically, one can choose to use an "unconstrained" model that does not force the mixtures to sum to 1 and allow fractions outside zero and one. Although when an unconstrained model is used, the mixtures may occur outside "real" bounds (e.g., less than zero reflectance), by analyzing a mixture this way one can determine whether the "best"

endmembers have been found. When endmembers are outside the range of 0-1, it may indicate a poor choice of endmembers, a missing endmember, or indicate that the endmember of interest is not a true endmember in the data space.
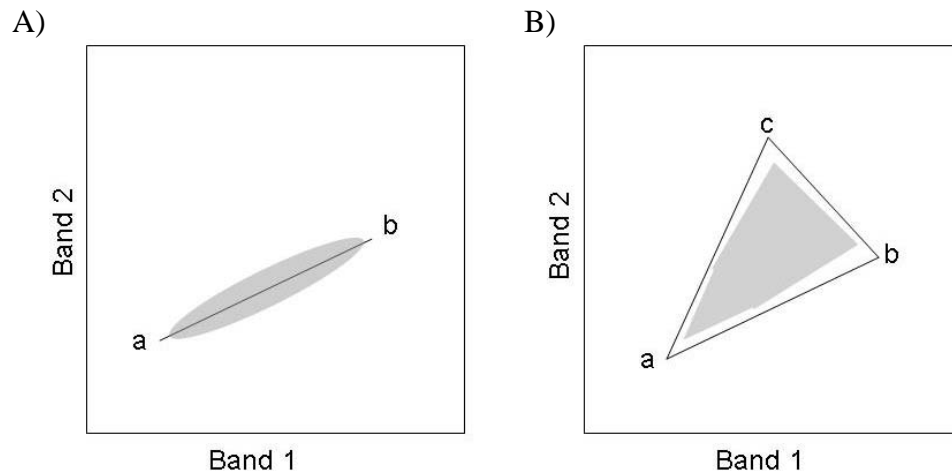
Figure 8: Linear mixture models. A) two-dimensional mixing (two endmembers). This model is equivalent to a physiological index, where the values range from zero at point a to 100%. canopy cover at point b. B) three-dimensional mixing (three endmembers), One axis in this model (a to b) represents from zero to maximum plant cover (or 100% soil to zero soil), and the axis (b to c) represents maximum or 100% green plant cover to maximum or 100% dead/senescent canopy cover, and the axis (c to a) represents maximum dry plant cover to zero plant cover (or 100% soil cover). The number of endmembers can be up to one more than the number of spectral bands although practical limitations allow only a few axes for any given data set.

Identifying the set of pixels on the outside of the data cloud (i.e., on the convex hull) is another method that can be used to determine how many and which endmembers are present in an image and to identify their spectra. One can then decide whether these spectra are relevant to include in an analysis based on the purpose of the user. The mixing model is extensible to higher dimensions and the number of mixing endmembers can be up to one more than the inherent dimensionality of the data. Practically however, due to similarity of the spectral endmembers (i.e., not fully orthogonal to each other) and the presence of noise in the data, the number of spectrally unique endmembers is significantly less than the inherent dimensionality. On a practical basis, the number of endmembers in images is probably half or fewer than the number of dimensions. Selection of too many endmembers will result in unstable solutions. Additionally, when the "best fit" composition of each pixel is calculated, the presence of endmembers in the equation that are not present in the pixel is a source of error (i.e., the method tries to fit a fractional abundance for all endmembers). Therefore it is best to keep the number to a minimum number of common classes.

4.3.6.1.   Practical Unmixing Methods

Two typical and very different forms of unmixing use 1) *a priori* known endmembers or 2) image derived endmembers. Using a set of known or assumed spectral endmembers for which the identity is known, can be used to derive the apparent fractional abundance of each endmember material in each pixel. These known endmembers can be calculated from (1) spectrometer data measured in the field, (2) obtained from a library of pure materials, or (3) estimated from models.

The mixing endmember matrix is made up of spectra from field measurements or a reference library. Unmixing is an over-determined linear least-squares problem. The mixing matrix is inverted and multiplied by the observed spectra to get a least-squares estimates of the unknown endmember abundance fractions. Constraints can be placed on the solutions to give positive fractions that sum to unity, as described above, or allowed to vary as mathematically determined. Shade and shadow are included either implicitly (fractions sum to 1 or less) or explicitly as an endmember (fractions sum to 1). While shade and shadow are not physical endmembers like a soil type of plant type, the reason for including them in an analysis is that they account for varying illumination in topographically variable conditions and some of the variation in the shadow component is related to canopy density and architecture.

The second unmixing method uses the hyperspectral image data itself to derive the mixing endmembers (Boardman and Kruse, 1994).   The inherent dimensionality of the image spectra is determined using a special orthogonalization procedure related to principal components and evaluated using other tools in the ENVI software program. This method was found to work very well with GPS located information for mapping invasive species at the scale of 3-4m pixels. It was relatively easy to find patches of one to several pixels in extent that could be identified on the imagery. This method was simpler and more effective than measuring field spectra and building a spectral library of materials for the analysis. This method only requires field personnel to GPS locate some specimens and does not require other more specialized training or equipment.

4.3.6.2.   Mixture-Tuned Match Filters (MTMF)

Mixture-Tuned Matched Filter™ (ENVI, Remote Systems, Inc.) is a hybrid method based on the combination of matched filtering and linear mixture theory (Boardman, 1998). This method combines the strength of the matched filter method with physical constraints imposed by mixing theory. Matched filtering is a technique adapted from electrical engineering that maximizes the target-to-background contrast by suppressing the response of the composite unknown background, thus matching the known signature (Chen and Reed, 1987; Stocker et al., 1990; Yu et al., 1993; Harsanyi and Chang, 1994). It provides a rapid means of detecting target spectra based on matches to specific library or image endmember spectra. This technique produces images similar to unmixing, but its advantage is that it does not require all endmembers to be known. It does, however, suffer from high false positive rates, where materials may be randomly matched if they are rare in a pixel (thus not contributing to the background covariance).

The equation for the MF image is as follows:

$$\Gamma = \frac{x^T \sum^{-1} d}{d^T \sum^{-1} d}$$ [4]

Where:
d = endmember spectrum
x = target spectrum
Σ = covariance matrix

Mixture tuning uses linear spectral mixing theory to constrain the result to feasible mixtures and reduce false alarm rates (Boardman, 1998). MTMF results are presented as two sets of images: 1) the MF score (matched filter image presented as gray-scale images with values from 0 to 1.0) which provides a means of estimating relative degree of match to the reference spectrum (where 1.0 is a perfect match); 2) an infeasibility image that indicates when mixing between the composite background and the target is not feasible. The best match to a target is obtained when the MF score is high (near 1) and the infeasibility score is low (near 0). The MTMF was one of the best methods found for identifying invasive plant species when there was little contrast between the invasive plant species and the background.

5. Results and Accomplishments

This study successfully mapped eight of the original species identified for this project: iceplant and jubata grass (Underwood et al., 2003, 2006a), fennel, giant cane (DiPietro et al., 2002), kudzu and lovegrass (Cheng et al., 2007), cheatgrass and knapweed (Noujdina and Ustin, 2007) the project set out to map. Two species were mapped with mixed success (tamarisk and common reed), and we were not able to analyze three species (Mediterranean grass, Asian mustard, and hydrilla) due to insufficient site conditions. The choice of analytical methods depended on the site conditions at the time of the airborne data acquisition. Figure 9 shows an example for mapping three invasive plant species, iceplant, jubata grass and bluegum, within two native plant communities, Coastal Sage Chaparral and Burton Mesa Mixed Chaparral at Vandenberg AFB based on an analysis that used the MNF method to classify species using field measured training data.
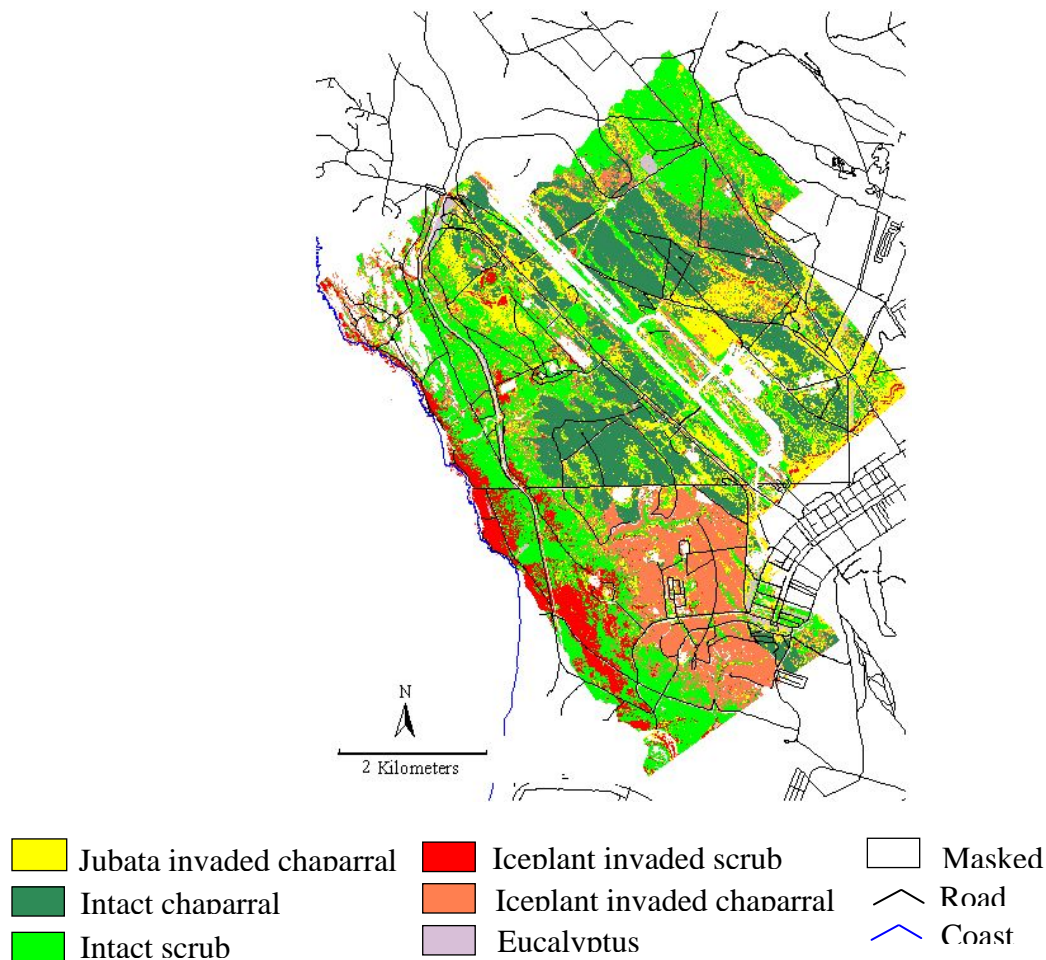
Figure 9: Classification of five community types using 14 bands from the Minimum Noise Fraction Transform performed on AVIRIS imagery (224 bands, 16m^2). Vandenberg Air Force Base, California (after Underwood et al. 2003; Underwood et al., 2006a)

In this case, the succulence of iceplant is distinctive against the xerophytic shrublands it is invading, jubata grass, as a large bushgrass, has a very different canopy structure than the shrub species, and the bluegum tree species has a different growth form than the shrub species.

Figure 10 provides a second example, from the eastern deciduous forest, mapping the invasive plant species, kudzu and lovegrass. In this example, physiological indexes were used to distinguish these invasive species from the native hardwood and pine forests. Lovegrass was detected primarily in open disturbed sites along roadsides while kudzu was located in forested sites and strongly influenced by site elevation.
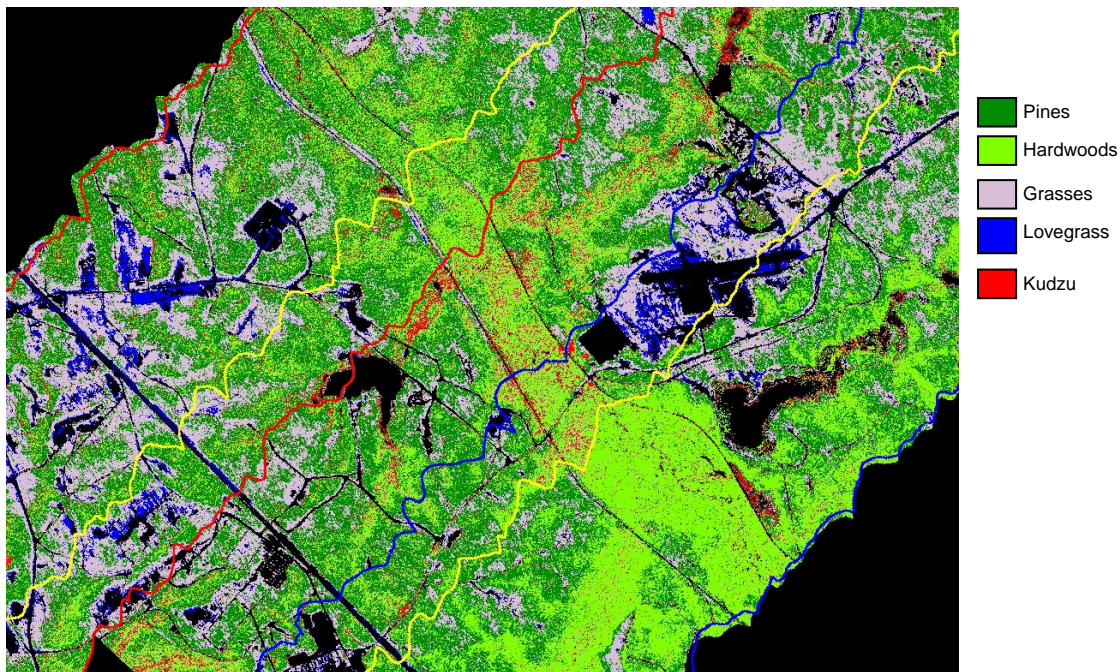
Figure 10: Mixed southern mixed pine-oak-hickory forest within Fort Benning, a National Heritage Site in southwestern Georgia. The classified image is a composite georeferenced mosaic of three AVIRIS flightlines (~6km width) shown by the yellow, red, and blue lines within the figure. Nine "physiological indexes" were used to create a classified map. Indexes were: water (NDWI, WBI), dry plant material (CAI, SIPI), and green foliage (EVI, MSR, MNLI, NDVI*SR, SAVI*SR).

5.1. Invasive Plant Species Mapping Accomplishments: Case Studies

The following five case studies were selected to provide examples of the range of results derived from this project and highlight their applicability to other conditions. They provide a more complete summary of the findings at each individual site. Additional details of other accomplishments are available in the published papers.

5.1.1. Case Study 1: Mapping Tamarisk (*Tamarix spp.*) Using Hyperspectral Image Data

Tamarisk is a phreatophytic invasive plant species invading rangelands throughout the western US. Tamarisk aggressively displaces native riparian species (Brotherson & Winkel 1986, Smith et al. 1998, Friederici 1995) and significantly alters hydrologic processes, wildfire frequency, and a wide range of other ecosystem properties.

Infestations of tamarisk cause a decrease in water tables, desiccating marshes, springs, and streams (Vitousek 1990, Vitousek 1986, Loope et al. 1988, Friederici 1995, Johnson 1987, Busch & Smith 1995, Di Tomaso 1998). High levels of water use result in accumulating salts at the surface and in the root zone, increasing soil salinity by up to 50-fold (Berry 1970). The roots and stems trap sediments, narrowing channels, thus increasing flow rates, contributing to both increased flood frequency and intensity (Di Tomaso 1998, Zavaleta 2000a, Zavaleta 2000b).

Previous remote sensing studies have demonstrated the potential for mapping tamarisk infestations using broad-band aerial photography (Everitt & Deloach 1990, Everitt et al. 1992, Everitt et al. 1996). These investigators found that tamarisk could be distinguished during periods when unique foliar colors are present, prior to leaf drop in late autumn. Such characteristics require images to be acquired in a brief time period in the year which may not be predictable, and may vary with environmental conditions. Therefore, image acquisition scheduling becomes very difficult or impossible and adding significant cost to the data collection. This study was undertaken to determine whether hyperspectral imagery could detect tamarisk independent of phenological conditions. At the continental scale, Morresette et al. (2006) used the MODIS weather satellite sensor at coarse pixel resolution (1km) to map potential tamarisk habitat for the United States using a logistic regression model trained on field observations. This approach is possible due to the association between water sources and tamarisk habitat and provides a basis for land managers to anticipate where an invasive species might spread, even when the direct detection of the invasive species is beyond the resolution of the instrument.

5.1.1.1. Materials and Methods

Three AVIRIS flightlines were acquired over the Yuma Proving Ground in southwestern Arizona on April 14th, 2002, during the driest year on record. Two of these flightlines were analyzed. The imagery had a nominal spatial resolution of 4 m x 4 m. Ground reference GPS and spectrometer data were collected in April 2002 and April 2004. Point data were collected for individual tamarisk shrubs and co-occurring native desert shrubs (palo verde, creosote, and ironwood), located in surrounding regions of irrigated agriculture and in the dry desert washes. These locations were selected to account for the variability in backgrounds where the target species could be found.

5.1.1.2. Hyperspectral Analysis

Images were atmospherically corrected to apparent surface reflectance using the radiative transfer package the Fast Line-of-sight Atmospheric Analysis of Spectral. Hypercubes (FLAASH[TM] developed by the U.S. Air Force Research Laboratory (AFRL) and the Spectral Information Technology Application Center (SITAC)). All image processing was performed in ENVI. Flightlines were georegistered and mosaicked together. We tested multiple analyses on the image data in order to determine the optimal method to detect tamarisk. Algorithms included Spectral Angle Mapper (SAM) classifications of the reflectance image trained with field spectrometer- and image-derived spectra, Maximum Likelihood (ML) classifications of the Minimum Noise Fraction (MNF) transformed image, ML classifications of physiological index images, linear spectral unmixing (SMA), and Mixture Tuned Matched Filter (MTMF).

From the suite of analysis, MTMF was the only technique that successfully detected tamarisk in the dry desert washes. This classification method may be especially useful in arid regions that are characterized by low plant cover and a high degree of mixed pixels with complex mosaics of various soil types. In arid regions, a major limitation for remote sensing is the high degree of spectral variability within species (Okin et al. 2001). As a result, the analysis was performed using several endmembers for each vegetation class. Where possible, endmembers were selected

from both the irrigated agricultural region and the wash region. Tamarisk was represented by the largest number of endmembers (n=5) to enhance classification accuracy, since tamarisk was the target species and its spectral characteristics had the highest within species variability. The remaining classes were represented with two spectra each. A map was generated from the MTMF results by assigning pixels to the class receiving the highest MTMF score above a threshold of 0.3 (Figure 11).

5.1.1.3. Results and Discussion

Tamarisk was the most spectrally variable class (Figure 11). Individuals in the area surrounding the irrigated agriculture were larger and had more vigorous growth and foliage than all native species and thus exhibited stronger "green" vegetation spectra. Irrigated tamarisk occurred in pure pixels that were not mixed with the soil background and, as a result, were clearly distinct from all native vegetation and identifiable by all classification methods. However, tamarisk also occurred in the dry desert washes, and in this setting tamarisk shrubs were of much smaller sizes and canopies, and supported significantly less green leaf area. Pixels were therefore mixed and tamarisk spectra overlapped significantly with the native species (Figure 11) which were also in a dormant or semi-dormant condition in late summer during a multi-year drought.
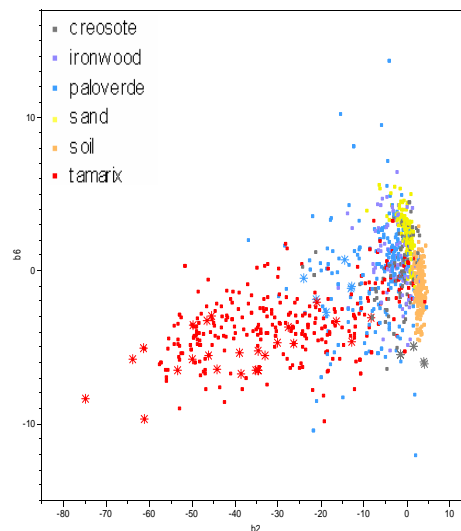
Figure 11: Field GPS located points on MNF bands 2 (corresponding to the gradient between vegetation and bare ground) and 6 (brightness in the infrared and/or depth of water absorptions. Asterisks are points collected in April 2002, squares were collected in April 2004.

Because tamarisk occurred in mixed pixels in the desert washes, the combined unmixing and matched filter approach, MTMF, was necessary for detection. The MTMF map of tamarisk is shown in Figure 12. This map detected 81.46% of known tamarisk (including 8 of the 19 wash individuals) with a user's accuracy of 91.16% (Table 5).
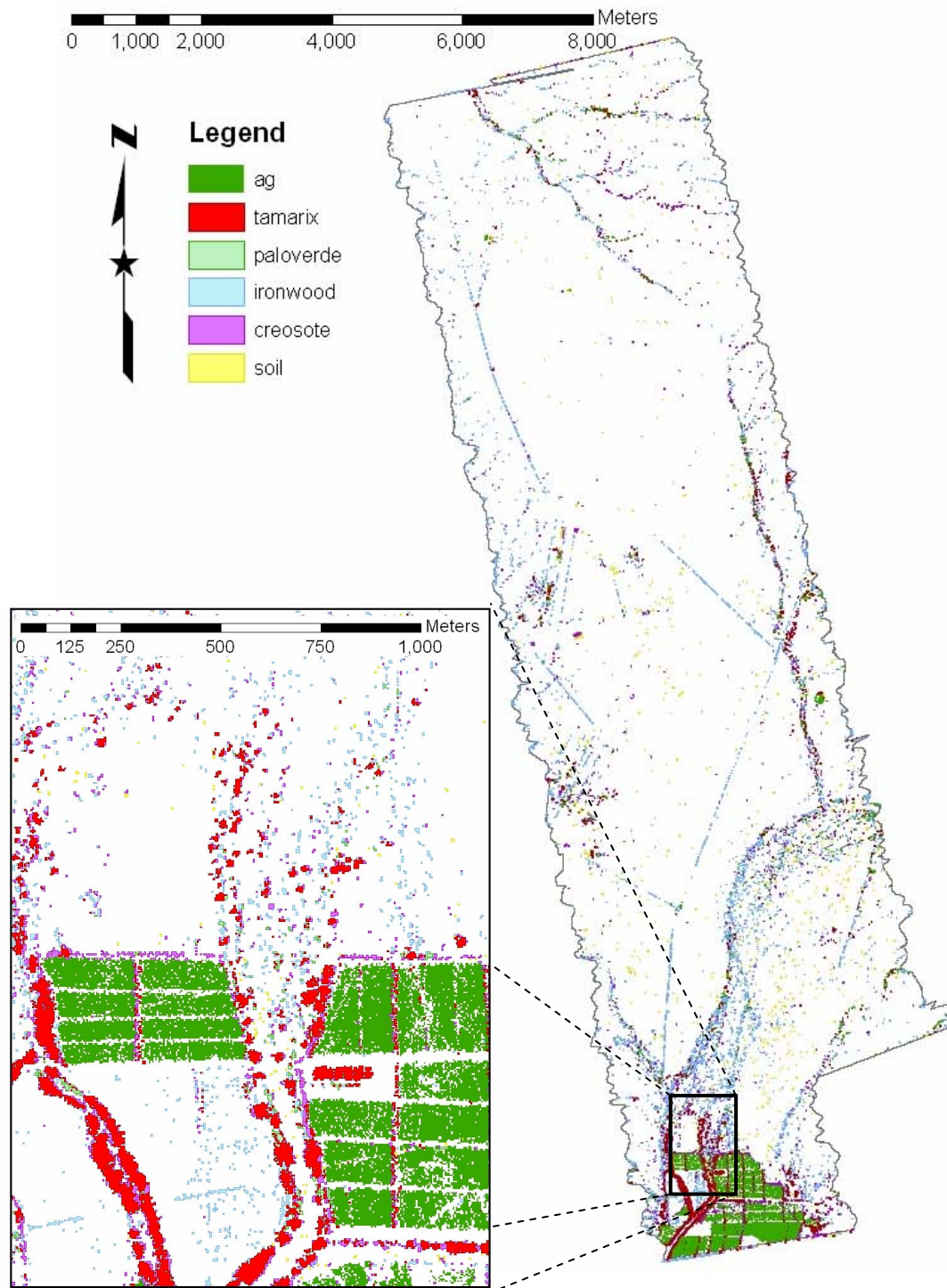
Figure 12: Tamarisk map generated from mixture tuned matched filter results. Pixels were assigned to the class with the highest MTMF score, if greater than 0.3. Unclassified pixels are white.

Table 5: Confusion matrix for MTMF tamarisk classification

		Map			Producer's accuracy
		Tamarisk	Other	Total	
Field	Tamarisk	268	61	329	0.81
	Other	26	410	436	0.94
	Total	294	471	765	
User's accuracy		0.91	0.87		
Overall accuracy		0.89			

These results demonstrate that it is possible to map tamarisk with hyperspectral image data, even under extreme drought conditions where it is dormant or severely stressed. Indeed, it is only possible to detect wash tamarisk with the detailed spectral information from hyperspectral image data combined with advanced image processing tools. Although tamarisk may have reduced ecosystem impacts where it occurs in dry washes, it is important to monitor it invasive plant species throughout the range of its infestation; these plants will grow opportunistically, and serve as sources for future spread.

5.1.1.4. Conclusions and Lessons Learned

Mapping invasive tamarisk, such as that next to agricultural fields, is among the easiest of detection problems because it occupies a particular zone within the habitat and can often tap water resources unavailable to other species in the habitat. The result is that tamarisk typically has a larger canopy and denser, healthier foliage than other species in the environment. Pixels of inappropriate habitat type can be masked to increase the likelihood of detection, and to further focus on the physiological differences between the native and invasive species. However, in the case of the dry washes occupied by dormant species, it becomes difficult to distinguish tamarisk. The detectable spectral signal comes from plant photosynthetic pigments. The strength of the spectral signal from dry plant material like woody stems and dry leaves is weak and can often be confused with spectral features from the soil, particularly in arid environments where different clay minerals, organic matter and carbonate may be present.

5.1.2. Case Study 2: Mapping Cheatgrass (*Bromus tectorum*) Using Multi-date AVIRIS Data

Cheatgrass is the most aggressive invasive plant species invading the shrub-steppe ecoregion of the Great Basin. The Yakima Training Center in eastern Washington is representative of this ecoregion type and has a major problem with management of cheatgrass. A mixture tuned matched filter (MTMF) algorithm was used to map percent cover of cheatgrass on a merged multi-seasonal dataset derived from two hyperspectral AVIRIS flightlines. The MTMF analysis utilized transformed bands derived from a minimum noise fraction (MNF) output.

5.1.2.1. Materials and Methods

AVIRIS hyperspectral images were acquired on July 27, 2000 and May 5, 2003. Spatial resolution of the 2000 and 2003 images were 3.8 m and 3.0 m, respectively. The datasets

consisted of the 224 contiguous spectral bands ranging from 400-2500 nm, at approximately 10 nm per channel. The two dates of data acquisition corresponded to two distinct life stages of cheatgrass: in May it is well established and vegetative whereas native vegetation is just emerging from winter dormancy. In July, the cheatgrass lifecycle is complete and it has senesced or is dormant while native grasses are still green. In both cases, spectral signals of the invasive plant species are distinct from that of the background vegetation. It was hypothesized that combining data from these two different growth states of cheatgrass would improve the contrast with the surrounding vegetation, increasing accuracy of the invasive plant species maps. The field data used to train the classifier and to build and test the models was collected in spring 2002, spring 2003, and summer 2005. Vegetation plots were sampled within the overlapping area of the two image dates. Data collected contained detailed descriptions of vegetation cover, including species names, percent cover, vegetation phenological stage, and type of soil where possible.

5.1.2.2.   Hyperspectral Analysis

Both hyperspectral datasets (July 27, 2000 and May 5, 2003) were preprocessed in ENVI software and atmospherically corrected using Atmospheric CORrection Now^{TM} (ACORN4, Imspec, Inc.). The resulting two 176-band, atmospherically corrected datasets were applied with the minimum noise fraction (MNF) rotation. MNF transformation was performed to reduce the dimensionality of the data and to prepare the input for the next step, mixture tuned matched filtering (MTMF). The first 22 MNF bands from each date where chosen as they contained the majority of the spectral variation. These were geometrically corrected and merged into a single dataset. Georectification was carefully executed to ensure that the datasets from different dates matched each other when combined. All three datasets (spring, summer and multi-date) were spatially trimmed to the aerial extent of the overlapping area. The ability of each dataset to detect cheatgrass using the MTMF algorithm was compared.

MTMF was chosen as the preferred algorithm because it incorporates the best properties of spectral mixture analysis (SMA) and matched filters (MF) and avoids the drawbacks of both. The advantage of the MTMF method is that it only requires knowledge of the target species endmembers. Pixels which have an MF score above the background value and a low Infeasibility index are those considered to have a high probability of being identified correctly. MTMF was run on the three datasets using training data of endmembers derived from pure dense cheatgrass plots, as determined from the field data.

Multiple linear regressions based on the MF score and Infeasibility score versus field-estimated percent cover yielded linear coefficients that were used to fine-tune the cheatgrass abundance maps in each case. Cheatgrass abundance was estimated for each pixel using equation 5 after Pontius et al. (2005):

$$Abundance = b1*Probability - b2*Infeasibility, \qquad [5]$$

Where:
   Probability – the probability output, MF score from MTMF image;
   Infeasibility – the infeasibility output from MTMF image
   b1 and *b2* are coefficients derived by regression analysis

5.1.2.3.   Results and Discussion

The comparison of correlation between cheatgrass abundance predicted by MTMF and the reference cover fraction of the invasive plant species for each of the three datasets demonstrated the following:

1.   Multi-temporal dataset produces a higher accuracy map of invasive cheatgrass abundance. The R^2 coefficient for the multi-temporal data is 0.79 (P = 3.88E-35), which is significantly higher than the R^2 of either single-date dataset: 0.41 (P = 3.63E-33) for July 2000 and 0.51 (P = 1.56E-29) for May 2003.

2.   Low spectral contrast between cheatgrass and the surrounding semiarid sagebrush-steppe vegetation in July resulted in a slight overestimation of cheatgrass abundance at low fractions. Certain species, such as tumble mustard (*Sisymbrium altissimum* L.) mirrors the spectral characteristics of cheatgrass in summer, thus lowering its detectability.

3.   Analysis of the May dataset resulted in overestimating cheatgrass abundance at low fractions and underestimating it at high cover fractions. This effect can be attributed to spectral confusion between cheatgrass and shrubs, which are both near full foliage and green at this time of year.

4.   The effects of over- and under-estimation of cheatgrass were reduced or eliminated when the multi-date dataset was used. The results based on the multi-date dataset provide better agreement with the field data.

The results of each MTMF analysis were reclassified to correspond to three abundance classes: high, moderate and low, and are shown in Figure 13. Quantitative accuracy assessment of the multi-temporal analysis is given in Table 6.
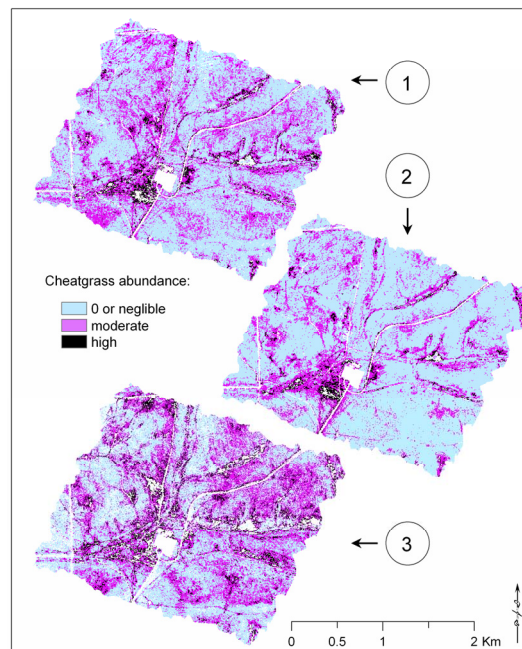

Figure 13: The results of MTMF classification for cheatgrass based on: (1) – multi-temporal spectral stack; (2) – July 2000 spectral data; (3) – May 2003 data

Table 6: Confusion matrix for multi-temporal data classification (Zstat = 15.88).

		M a p			Producer's accuracy
		Cheatgrass	Other	Total	
Field	Cheatgrass	92	17	109	0.84
	Other	26	91	117	0.78
	Total	118	108	226	
User's accuracy		0.78	0.84		
Overall accuracy		0.81			

Overall accuracy for the multi-temporal data classification was significantly higher then either of single date dataset: 81% vs. 70% (July) and 72% (May). Analyzing the multi-temporal dataset primarily reduced commission errors of cheatgrass detection: user's accuracies were 0.66, 0.67, and 0.78 in the July, May, and combined datasets, respectively. Omission errors were also reduced using the combined dataset; producer's accuracies were 0.76, 0.83, and 0.84 in the July, May, and multi-temporal analyses.

Further analyses of surrounding vegetation which weakened the cheatgrass abundance predictions in single season's images supported the advantage of using multi-temporal dataset when detecting invasive cheatgrass in an arid environment. The difference in cheatgrass detection using hyperspectral data acquired in different seasons corresponded to the seasonal phenologic stages of vegetation in the area. When summer hyperspectral data were used, cheatgrass was not detected in sparsely vegetated, arid areas because the spectral signal of dry vegetation is less distinguishable between individual species. This result is similar to tamarisk in the previous example. In contrast, when spring hyperspectral data was analyzed to detect cheatgrass abundance, it was overestimated in sparsely vegetated areas since spectral contrast between vegetation and soil is much stronger than that between different grass species. In the areas with mixed grass cover, cheatgrass was detected at a 30% threshold. These discrepancies are subdued when a multitemporal data set is used.

5.1.2.4. Conclusions and Lessons Learned

These results support the use of multi-temporal data to increase the accuracy of cheatgrass detection in the semi-arid rangelands of the western states. Spectral detection of rangeland species depends on additional structural differences between the density of the canopy and/or foliage, the canopy architecture or branching pattern characteristic of the species, and the biochemical composition of the species. Detection requires that some elements are different between the invasive species and the native species. The simplest is plants having different growth forms e.g., shrubs and herbaceous/grass species. The spectral contrast between cheatgrass and the surrounding environment is controlled by their respective phenologies, which are captured when measured at multiple times. It is more effective when each measurement is timed to specific phenological stages or events in the life cycle of the target species. It is likely that multitemporal data from different seasons will enhance the detection of most invasive species.

5.1.3.   Case Study 3: Mapping Common Reed (*Phragmites australis*) Using Hyperspectral Image Data

Phragmites australis, or common reed, has a circumpolar distribution, wide salinity tolerance, and a broader elevation range than most wetland plants. It has been described as the "most widely distributed angiosperm" (Mal & Narine 2004). With its worldwide distribution, its origin is unclear, however the species or possibly a subspecies is also considered to be native to the U.S. (http://plants.usda.gov/), making its listing as an aggressive invasive plant species unusual. Peat core data indicate that it has been a part of the mixed tidal wetland plant communities of North America for the past 3,000 years (Niering et al. 1977, Orson et al. 1987). Beginning in the 19th century and accelerating over the past century, the distribution of common reed has undergone conspicuous changes as it expanded its range and abundance by invading fresh and brackish wetlands. Once a rare, benign component of North American wetlands, it is now aggressively spreading and has come to dominate many of these systems with negative impacts on ecosystem processes and native wildlife (Marks et al. 1994, Meyerson et al. 2000). Here common reed was mapped in hyperspectral imagery on two islands of Aberdeen Proving Ground in Maryland.

5.1.3.1.   Materials and Methods

Four AVIRIS flightlines were obtained over Aberdeen Proving Ground in eastern Maryland on September 12, 2002. The imagery has a pixel resolution of 4.3 m x 4.3 m (18.5m^2). Two of the four flightlines covering Carroll and Spesutie Islands were analyzed for common reed detection. Ground reference data were collected in August 2002 and November 2003. Point data were collected of common reed and co-occurring wetland and riparian species, including grass, trees, cattails, and sedges.

The flightlines were radiometrically calibrated and atmospherically corrected using Atmospheric CORrection Now^{TM} (ACORN4, Mode 1, Imspec, Inc.) to apparent surface reflectance. Both flightlines were registered to georeferenced aerial photos. Because the flightlines did not overlap, they could not be cross-calibrated to reflectance and therefore, were analyzed separately. Many analyses were tested on the image data in order to determine the optimal method of detection. Algorithms investigated included maximum likelihood (ML) classification and spectral angle mapper (SAM) classifier, both performed on the full reflectance data set. Additionally several data reduction steps were tried including: 1) the minimum noise fraction transformation (MNF), a two step principal components transformation of the original bands, and 2) calculating 19 physiological indexes, which are primarily ratios of band combinations. This was followed by an ML classification of physiological index images to species type, and a newer merged method termed mixture tuned matched filter (MTMF).

Different techniques produced the best map of common reed on each island. On Carroll Island, a SAM classification performed on physiological indexes optimized common reed detection. The 19 physiological indexes listed in Table 7 were calculated from the reflectance image, registered, and stacked into a single multiband image. Two groups of highly correlated indexes were found. The first group contained pigment indexes and indexes expected to correlate with the greenness and vigor of a plant: NDVI, mNDVI, SIPI, PI2, and NDNI. Additionally, CAI was highly correlated to NDNI. The second group included SGR and CAI. Three indexes contained little

information and excluded from further analyses: WBI, NPCI, and NDLI. To determine which set of indexes resulted in the best classification, each independent index was iteratively added and removed as classification inputs to determine their influence on classification accuracy. Finally, indexes were substituted with highly correlated counterparts to determine their effect on mapping. A SAM classification was performed on the final set of indexes chosen.

In contrast, none of the classification methods performed well on Spesutie Island (Table 8). The unmixing algorithm, MTMF performed on the first 16 MNF bands, however, did successfully map common reed. Training spectra were derived from the most spectrally extreme (the convex hull of the data volume) ground reference pixels in each class, and thus represented the purest, least mixed points. Pixels were assigned to the class for which they received the highest MF score, if greater than a threshold of 0.3.

5.1.3.2. Results and Discussion

Only 3 indexes were necessary to accurately map common reed on Carroll Island: NDVI, NDWI, and PRI, indicating that greenness, water content, and xanthophyll content provide sufficient information necessary to distinguish common reed from co-occurring species. The resultant SAM classification is shown in Figure 14. Overall accuracy was 80% based on the validation data (independent of the training data), and common reed was detected with producer's and user's accuracies of 71% and 74% (Table 7). Clearly the coherent spatial distribution pattern of common reed in the map (Figure 14) is consistent with expectations for a good classification (clumped distributions with no speckle).

Table 7: Accuracy assessment of SAM classification of NDVI, NDWI, and PRI on Carroll Island, including Spesutie Scirpus and cattail endmembers.

		M a p			Producer's accuracy
		Phragmites	Other	Total	
F i e l d	Phragmites	20	8	28	0.71
	Other	7	39	46	0.85
	Total	27	47	74	
User's accuracy		0.74	0.83		
Overall accuracy		0.80			

For all Carroll Island classifications, SAM outperformed the maximum likelihood algorithm, suggesting that the field data collected and training data used were highly characteristic of the entire common reed infestation on this island. Since relatively few endmembers were necessary per class, this also suggests that there was relatively little within class spectral variability and that each type occurred as large, discrete entities with relatively little mixing. This finding is confirmed by the poorer performance of spectral unmixing with MTMF analyses.
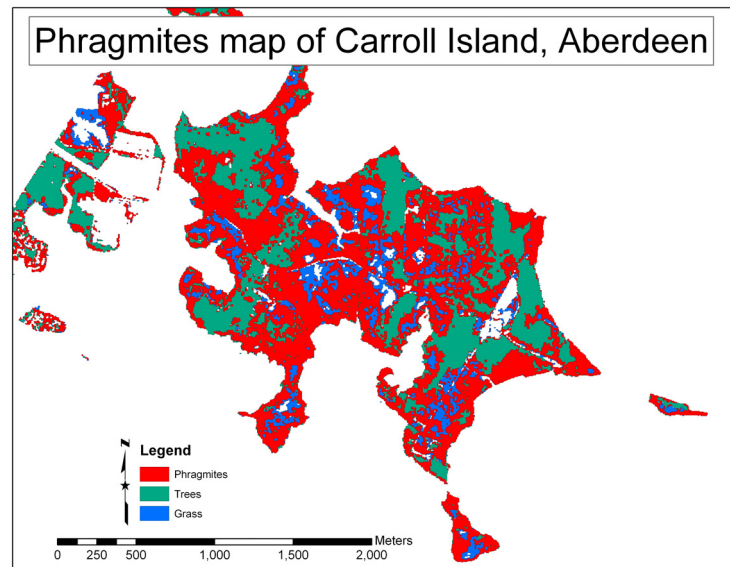
Figure 14: Spectral angle mapper classification of common reed performed on the physiological indexes NDVI, NDWI, and PRI.

The use of the MTMF algorithm was necessary to map common reed on Spesutie Island (Figure 15), which was supported by field observations that considerable mixing of wetland species occurs on this island. Despite this, the common reed map accuracy was exceptional using the MTMF, with an overall accuracy of 97% correct. Furthermore, 88% of known common reed was detected with a user's accuracy of 72% (Table 8). Common reed was most often confused with cattails.

Table 8: Accuracy assessment of MTMF analysis of Spesutie Island

		M a p			Producer's accuracy
		Phragmites	Other	Total	
F i e l d	Phragmites	173	23	196	0.88
	Other	68	2553	2621	0.97
	Total	241	2576	2817	
User's accuracy		0.72	0.99		
Overall accuracy		0.97			

5.1.3.3.   Analogies to Giant Reed (*Arundo donax*)

Both giant reed and common reed are tall, cane or bamboo-like grass species. Giant reed grows 6-10m tall in contrast to the 2-6m height of common reed, and is more characteristically found in freshwater wetlands and riparian zones. However, both form dense impenetrable stands, which can rapidly extend by vegetative propagation. Giant reed, introduced to California from the Mediterranean has become a serious pest throughout the streams and rivers of coastal California and the Central Valley. Unlike other wetland species, giant reed significantly increases the probability of wildfire in riparian zones, due to its accumulated biomass and because it is dry or senescent in the summer drought high fire risk period in the West.

Because of the similarity of their growth forms and the habitats they invade, a comparison of the techniques to map these similar species is informative. Spectral indexes and MTMF successfully detected common reed under different background conditions. Indexes were not systematically applied to map giant reed (it was one of the first SERDP sites at the project worked at), however the depth of the water absorption feature at 970 nm successfully isolated giant reed from other grass and chaparral species. This absorption feature is the basis of the WBI index and is a metric of the amount of foliar water present in a pixel. It is often highly correlated to NDWI, which contributed to common reed detection on Carroll Island. By itself, the water absorption feature, however, did not consistently separate giant reed from aquatic vegetation species, which also had high water contents. It is possible that if this feature had been used in combination with the suite of 19 vegetation and pigment indexes, such as NDVI and PRI that were successful in mapping common reed on Caroll Island, Aberdeen Proving Ground, giant reed could have been successfully mapped with spectral physiological information. Clearly giant reed has structural distribution features that suggest this method would be successful: large, relatively uniform monotypic clumps that are continuous over many pixels. Both giant reed and common reed case studies highlight the importance of foliar water content in successful mapping of invasive plant species (as does unpublished data [not shown] for cheatgrass) but suggest that additional information about plant vigor and pigmentation are also necessary to fully characterize the species.
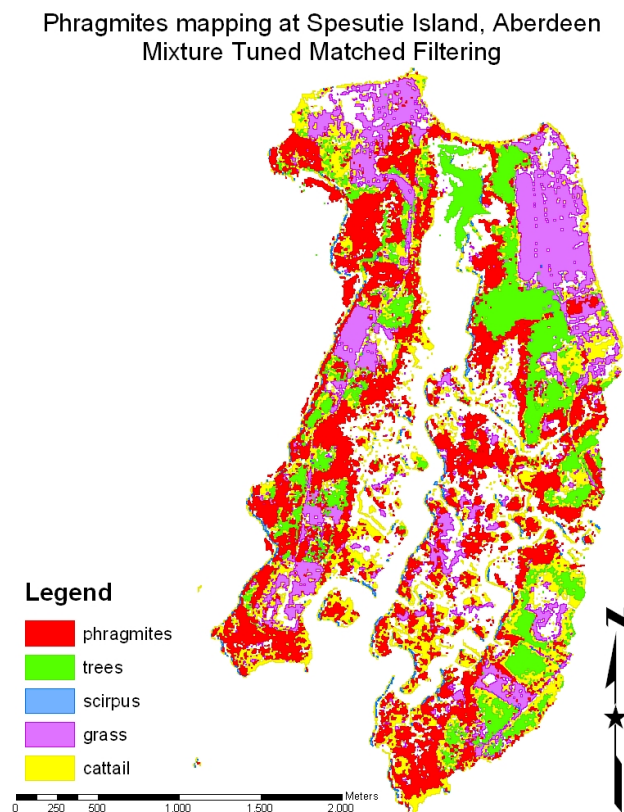

Phragmites mapping at Spesutie Island, Aberdeen
Mixture Tuned Matched Filtering

Figure 15: MTMF map of common reed (mapping threshold = 0.3) for Spesutie Island.

37

5.1.3.4.   Conclusions and Lessons Learned

Again, the success in spectral detection is dependent on the variability of the invasive species pixels versus the other classes in the data set. If all locations for the invasive species are consistent and the range of variability is small, the likelihood of detection is reasonably good. However, if the invasive species class overlaps with other classes, it becomes difficult to consistently detect them. Also, if the background vegetation is heterogeneous, this decreases the potential for detecting the invasive plant species in its presence. This example for increased difficulty in mapping common reed on Spesutie Island is consistent with the findings shown for tamarisk in the dry washes and for iceplant in the Burton Mesa Mixed Chaparral.

5.1.4.   Extending Mapping Beyond the Image Space Using CART Modeling

5.1.4.1.   Introduction

As shown in previous sections remote sensing provides a very useful tool for mapping plant communities and invasive species. However, for larger facilities, it could be expensive and difficult to acquire aerial imagery for the entire base. Therefore, a technique that allows base managers to extend mapping capabilities beyond the boundaries of the actual image itself would be useful. As part of the SERDP project, the hyperspectral imagery at VAFB was used to create models which would predict the locations of plant communities and invasive plant species. The technique found most useful was classification and regression tree (CART) analysis.

CART trees explain the variation of a single response variable through repeated partitioning of a dataset into increasingly homogeneous groups using combinations of categorical or numeric explanatory variables. Trees are represented graphically beginning with the complete data set at the root node. Each binary split separates samples at the next lower branch as defined by a value or range of values for the explanatory variables. Splitting continues until the tree is over-fitted to the data and the final groups occupy the leaves or terminal nodes of the diagram. The over-fitted tree is then pruned to the desired size based on complexity plots and a consideration of the research goals.

5.1.4.2.   Mapping Vegetation Community at VAFB

The primary reason that VAFB was chosen for these modeling studies is that the project had the most detailed field data for this site both from CSTARS and the project's industry partner, Dynamac Corporation. All field data were compiled and GIS layers available for VAFB were used in a CART analysis. In total, 579 field locations were used. The field data provided the response variable (community type) and the GIS layers provided the explanatory variable. Some layers that were used are shown in Figure 16. The major two communities identified at VAFB were the Southern Coastal Sage Scrub and Burton Mesa Maritime Chaparral communities.

The CART model provided a list of rules defining the important splits where the scrub and chaparral were located on the CART tree (see Figure 17). The tree was used and it was applied to the GIS layers for the entire study area to develop a map of the scrub and chaparral plant communities. To provide a qualitative check, the CART map was compared with an existing

vegetation map provided by VAFB. As shown in Figure 18, there is strong visual agreement between the two maps.

To provide a more rigorous quantitative evaluation, the CART map was compared to the map developed from the hyperspectral imagery. Then the community information was extracted from the vegetation image categories and produced a map which divides the image area into scrub, chaparral and "other".   Using the hyperspectral map as a baseline, it was determined that the CART model predicted the location of the vegetation communities within the image area with 76.9% accuracy (see Figure 19). This is excellent accuracy showing that the CART model is robust and a powerful tool for plant community mapping.

5.1.4.3.   Identifying Factors which Promote Invasion

Another avenue that was explored with the VAFB data set was whether the environmental and disturbance factors that would promote the spread of invasive plant species at VAFB could be identified. This analysis produced promising results modeling the invasive iceplant and jubata grass at VAFB.

One of the problems with modeling invasive plant species is that the environmental (natural "steady state") and disturbance factors are usually confounded making them difficult to tease apart. From an environmental management standpoint it would be most useful to know which areas are especially vulnerable to invasion due to environmental conditions and should be monitored more closely. And whether these factors could separately determine which disturbance factors promote invasion, the base personnel could then make informed decisions about whether to allow a specific type of disturbance in the vulnerable areas.

To model a site in this degree of detail requires an exceptional number of samples and would be near impossible using field data. However, by using the hyperspectral map as the data set, produced in effect, 100% sampling within the image area. Since there were 1.65 million classified pixels within the image, the data could be subset many times to provide test data, verification data and validation data.

To separate the natural environmental conditions and disturbance factors required a two-stage approach:
1. Identify the most favorable environments for invasive plant species growth
2. Identify the most important disturbance factors influencing invasive plant species growth in both favorable and unfavorable areas.
3.

5.1.4.4.   Analysis of Environmental Factors

To identify the location of the most favorable environments for iceplant and jubata grass, the work began with an assumption that invasive plant species will cluster most densely where environmental conditions are optimal and sparsely where conditions are unfavorable. Using this assumption, the first step of the analysis was to locate areas were there were high invasive plant species densities and areas where invasive plant species densities were low or not present. This gave each pixel a score representing the number of neighboring pixels were also invaded. The

pixels with the highest score were considered the densest clusters and thereby, the optimal environments. Figure 20 shows the location of the optimal environments for iceplant and jubata grass.

Then a CART analysis was performed on the best and worst environments for invasive plant species to determine associated characteristics. The results are described for iceplant in Table 9 and jubata grass in Table 10. The CART analysis shows that iceplant invasion of the chaparral community seems to be most favorable when iceplant is located in hilly areas having deep soils that are prone to water erosion. Iceplant does not do well in riparian zones or areas too far from the coastline.

For the scrub community, iceplant grows best in hilly terrain with deep soils that are prone to erosion. However, the slopes most favorable to iceplant invasion are much steeper (>7%) and soils are much more erodable (rank >5). Iceplant has an affinity for acidic soils with a high storie index. It does not grow well in riparian zones (although it is less limited than in the chaparral community), in soils that retain water or on north and west facing slopes.

The results of the jubata grass analysis are summarized in Table 10. Jubata grass invasion is favored in areas with markedly different conditions than iceplant. The factors favorable for jubata grass invasion include flat areas away from the coastline, with stable sandy soils and low shrink-swell (clay) potential and low erosion potential. In contrast to iceplant, jubata grass does well on north and west facing slopes.

5.1.4.5.   Influence of Disturbance Factors

The above classifications were used to determine which areas of the base are most at risk for plant species invasion. The next step in the analysis was to explore how the high risk (favorable) and low risk (unfavorable) areas respond to disturbance factors (Tables 11-13). The project sampled both the high risk and low risk areas separately and looked at which disturbance factors were important in each environment.

The CART results are rather complicated, but overall trends can be extracted for scrub-iceplant, chaparral-iceplant, and chaparral-jubata grass as shown in Tables 11, 12, and 13, respectively.

Proximity to fences enhances the spread of iceplant and yet seems to provide a deterrent to the spread of jubata grass. Proximity to firebreaks and parking lots enhances both jubata grass and iceplant spread. Proximity to roads enhances the spread of both invasive plant species and yet this positive feedback seems to be ameliorated somewhat if the roads are paved. The proximity to hydrologic structures can either enhance or deter spread of invasive plant species. The mixed results may be due to different types of hydrologic structures. Some structures enhance drainage (ditches) and others enhance soil moisture (dams). Dams in particular seem to favor the spread of jubata grass. The other major factor affecting invasive plant species spread is fire. Jubata grass seems particularly sensitive to fire and is deterred even in favorable environments if there has been a recent burn. Iceplant, on the other hand, was not particularly sensitive to fire and in favorable environments, and recent fires seemed to enhance the spread of iceplant.

Table 9: Summary of environmental factors for iceplant invasion from CART Analysis.

Community	Favorable factors for iceplant	Unfavorable factors
Chaparral – Iceplant	Distance to coast between 2284 and 2982 m Hilly terrain (slope > 1%) Storie Index between 24 and 30 Deep soils (depth to bedrock > 4.1m) High TRMI > 18	Closeness to fresh water (< 570 m) Non-erodable soils
Scrub - Iceplant	Nearness to coast Hilly terrain (slope > 7%) Storie Index > 34 Deep soils (depth to bedrock > 4.1m) Highly erodable soils Acidic soils (minimum pH <= 5.1)	Nearness to fresh water (< 172m) Moist soils North and West facing slopes (Aspect > 276)

Table 10: Summary of environmental factors for jubata invasion from CART Analysis.

Community	Favorable factors for jubata	Unfavorable factors
Chaparral - Jubata	Distance to coast > 1619m Sandy soils Storie Index > 34 Storie Index < 24 Flat terrain (slope < 2%) Distance to fresh water > 105m North and West facing slopes (Aspect > 290) Soil class ML High TRMI > 24	Soil class SM Shrink-swell >1 Erodable soils

Table 11: Summary of disturbance factors determined to be important for the spread of iceplant in scrub community.

Community	Favorable Factors for Invasive Plant Species Spread	Unfavorable Factors for Invasive Plant Species Spread
Scrub - Iceplant Unfavorable Environment	Fences Firebreaks Multiple lane roads	Paved roads Hydrologic structures
Scrub - Iceplant Favorable Env.	Fences Building Parking lots Hydrologic structures Unpaved roads	

Table 12: Summary of disturbance factors determined to be important for the spread of iceplant in chaparral.

Community	Favorable Factors for Invasive Plant Species Spread	Unfavorable Factors for Invasive Plant Species Spread
Chaparral - Iceplant Unfavorable Env.	None identified	None identified
Chaparral - Iceplant Favorable Env.	Recent fires (within ~ 30 yrs) Hydrologic structures Fire breaks Buildings Parking lots	Hydro structures

Table 13: Summary of disturbance factors determined to be important for the spread of invasive plant species

Community	Favorable Factors for Invasive Plant Species Spread	Unfavorable Factors for Invasive Plant Species Spread
Chaparral - Jubata Unfavorable Env.	Fire breaks Roads (particularly unpaved) Hydrologic structures Parking lots	Recent fires (within ~ 30 yrs) Fences
Chaparral - Jubata Favorable Env.	Fire breaks Parking lots Roads (particularly unpaved) Dams	Recent fires (within 3 yrs) Fences
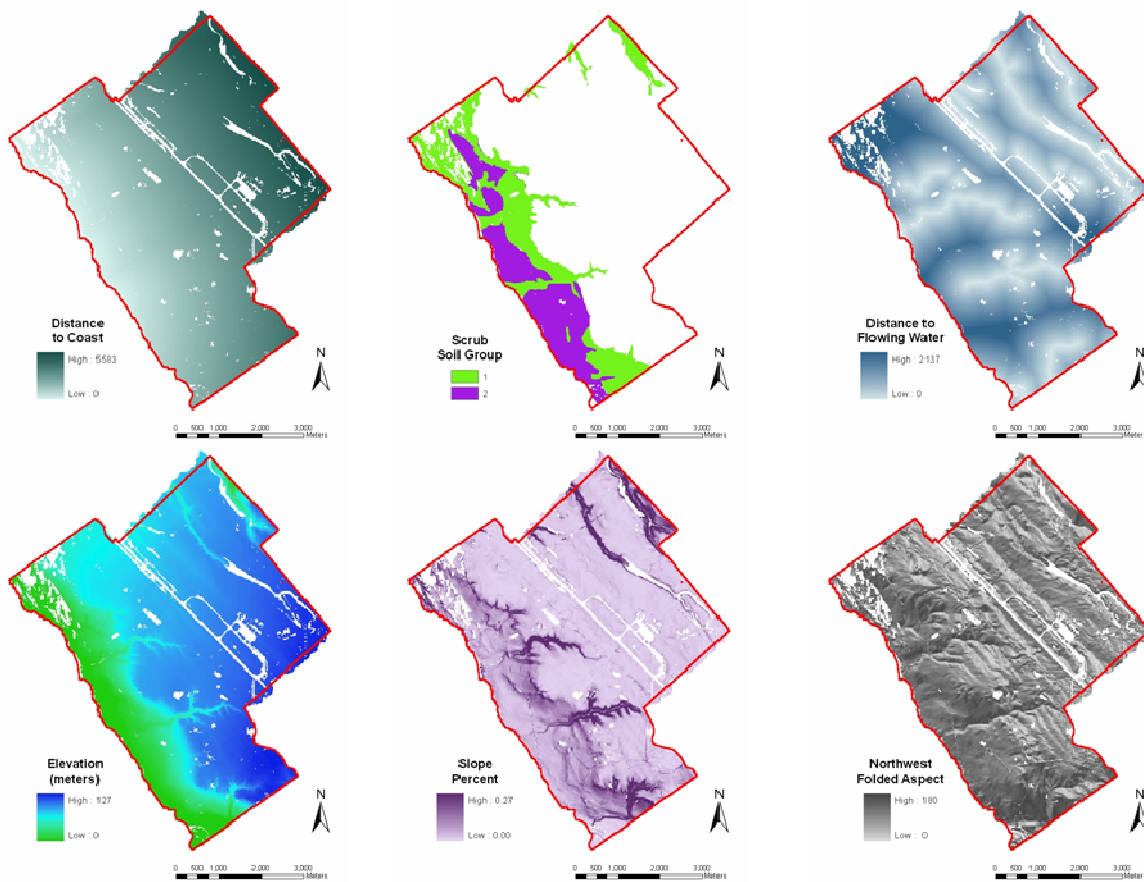
Figure 16: Example of GIS layers used in the CART analysis.
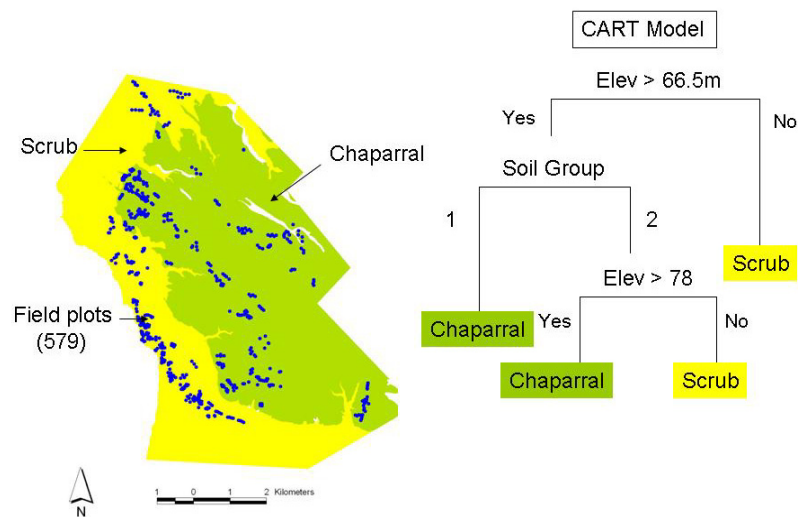
Figure 17: CART model for scrub and chaparral community mapping. Note area is larger than original AVIRIS flightlines.

CART Community Map   Existing vegetation map
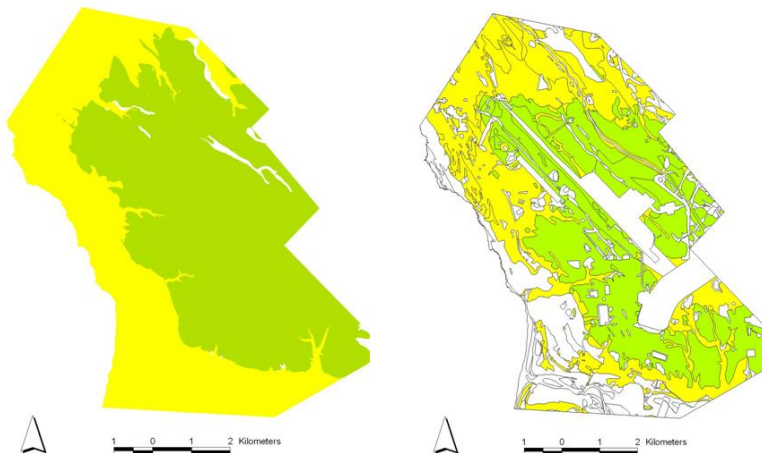
Figure 18: Comparison of CART map with VAFB preexisting vegetation map.
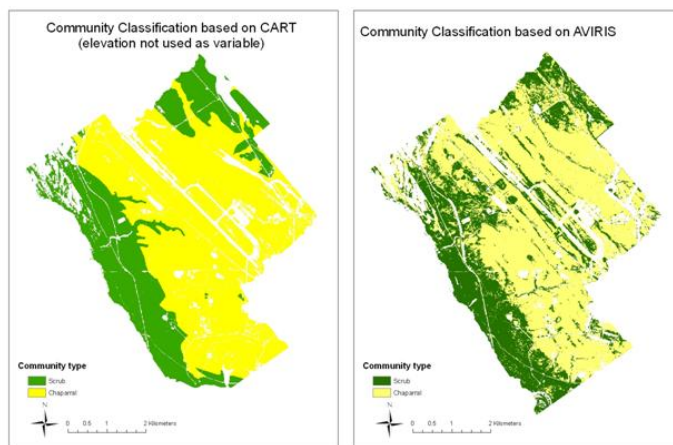
Figure 19: Comparison of CART vegetation community map with classified hyperspectral map.

		AVIRIS		
		Scrub	Chaparral	Total
CART (without elevation)	Scrub	344,858	161,320	506,178
	Chaparral	191,657	833,484	1,025,141
	Total	536,515	994,804	1,531,319
	Accuracy	64.3%	83.8%	76.9%

Figure 20: Optimal locations for invasive plant species predicted from CART model.

5.1.4.6.   Conclusions and Lessons Learned

Because it seems likely that hyperspectral imagery will not be able to identify the earliest establishment of an invasive plant species in a new site, modeling is likely to be the best method for predicting sites of invasion and the most vulnerable locations. There are many modeling methods that have potential for site-specific modeling. This study i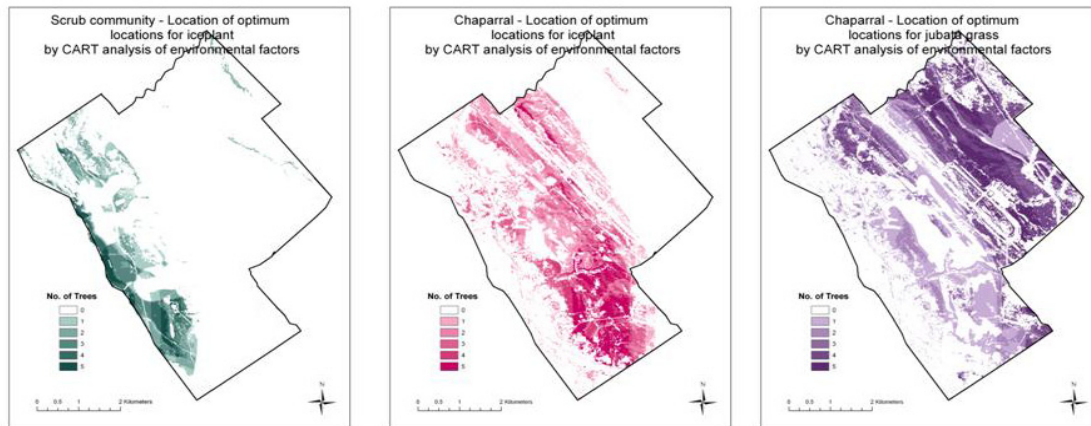nvestigated CART modeling as it is relatively simple binary tree to develop, assuming a GIS database is available and can be intuitively interpreted. It can incorporate various types of data and because variables can be incorporated at various levels in the model, provides a basis for handling non-linear relationships among variables.   CART modeling provided significant insights into factors that promote the spread of invasive iceplant from natural environmental conditions and from human activity. This type of information would be useful to resource managers for understanding where invasions might spread and for anticipating the impact of activities on the spread of invasive plant species.

5.1.5.   Tree Decline at Fort Benning

5.1.5.1.   Introduction

A small subproject was attempted in the final year of the study at Ft. Benning, to determine whether hyperspectral AVIRIS data could detect loblolly pine tree decline as noted in the field by base biologists and others. A total of 38 tree plots with GPS locations were provided to us by Dr. Lori Eckhardt of Auburn University. The data set contained 22 healthy trees and 16 trees with evidence of Loblolly Pine Decline (LPD). We used this data in two ways:

1) We performed a preliminary review to evaluate whether hyperspectral data collected at Fort Benning could be used to map trees with LPD. The plots in Dr. Eckhardt's study were used as ground control points.

2) Performed a CART model with Ft. Benning GIS data to see if local environmental parameters correlate with LPD.

5.1.5.2. Results of LPD Derived from Hyperspectral Mapping

The location of Dr. Eckhardt's plots in relationship to our AVIRIS-classified map is shown in Figure 21.
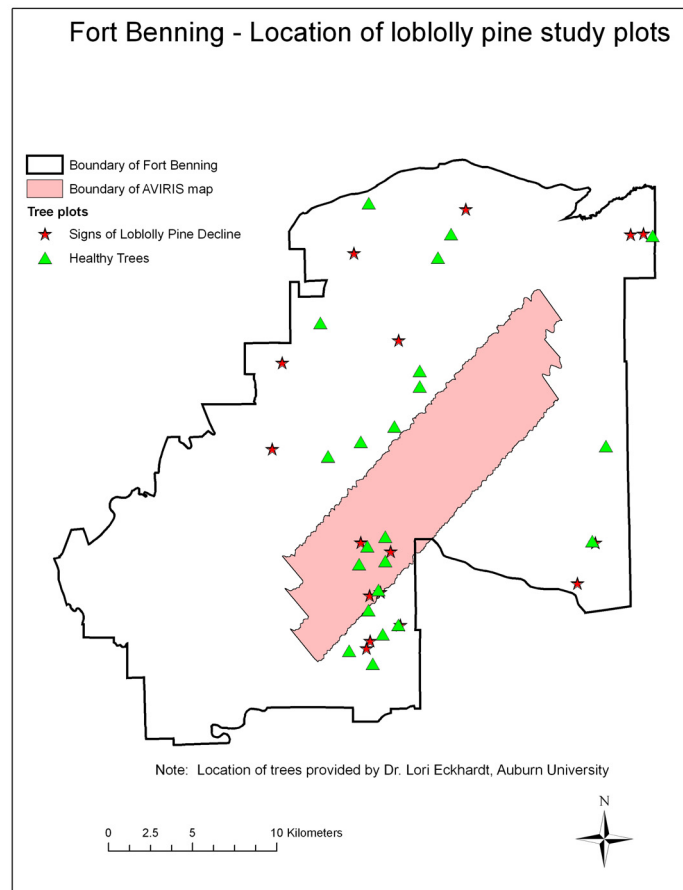
Figure 21: Location of loblolly pine study plots at Fort Benning

There were only 5 healthy trees and 4 trees with LPD within our flightlines. Therefore, there was insufficient training data to directly map the LPD using the hyperspectral signatures. We tried several methods e.g., PPI, MTMF, and SAM methods without clear success. There are several likely explanations why these results were poor:

1) There were insufficient samples within the AVIRIS flightlines to make a valid statistical comparison of the spectral signatures of the healthy vs. diseased trees.
2) We had difficulty co-registering the precise location of the trees to our hyperspectral imagery. Normally when we register vegetation targets on the ground, we try to find plots that encompass several pixels within the image, thus reducing the potential that the field data will be in a different pixel due to georegistration error. Since the pixels in the Fort Benning imagery are 3.5 m x 3.5 m in size (~ 12 m^2), we typically try to locate targets of at least 50 sq meters.
3) Trees with LPD were scattered within the distribution of healthy trees, therefore location errors and no clear spectral signal made it impossible to adequately train the classifier.

4) The proximity of healthy trees to trees with LPD, resulted in few pixels containing only trees with LPD.

5) There was a time lag of 2 to 3 years between the time that the hyperspectral data was acquired and the tree plots were measured. The progress of the disease may have significantly changed during this interval.

6) The possibility that other analysis methods would have detected the LPD in the hyperspectral data. We did not exhaustively test a wide range of possible algorithms.

A new feasibility study is planned to collect high spatial resolution hyperspectral data in an area covering known LPD and healthy trees. However, from this study, as we discuss in the next section, we were still able to use the hyperspectral imagery in combination with a CART model to map potential LPD sites indirectly.

5.1.5.3. Results of Hyperspectral Mapping of Pine Species

Pine endmembers were used in a tree classification using the Spectral Angle Mapper method. The results for tree type mapping were much better than tree health. The pixels classified as pine are shown in Figure 22. The pine endmembers included pure pine (2,107,899 pixels), pine mixed with hardwoods (195, 781 pixels) and pine mixed with grasses (608, 926 pixels). This is 10 % of the total area within the three flightlines (364,173,732 m^2).
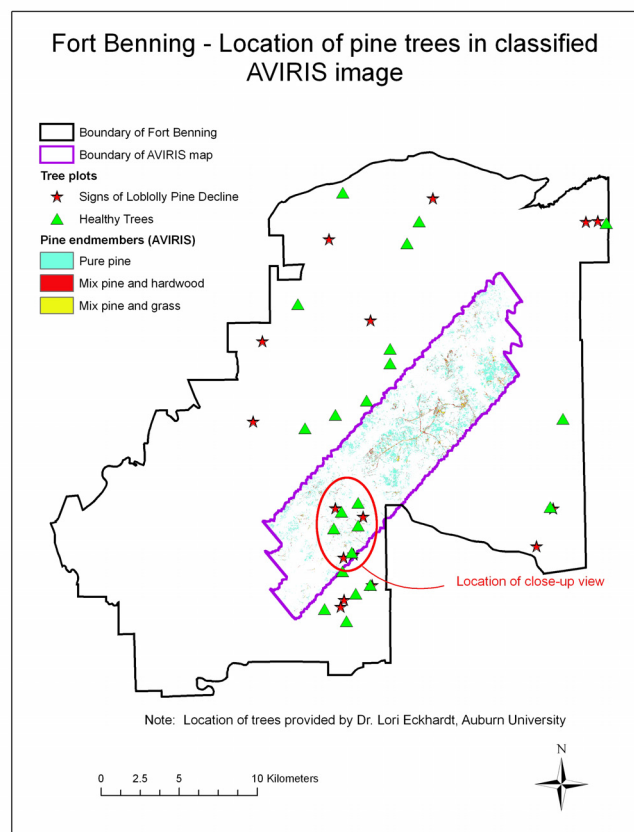
Figure 22: Location of pine trees in classified AVIRIS image

A close-up of the area where the flightlines overlap with Dr. Eckhardt's plots is shown in Figure 23. It is difficult to compare the plots directly with the classification, due to uncertainties in the co-registration.

Figure 23: Close-up of overlap of flightlines and plots

5.1.5.4. Results of CART Models

We used the GIS from Ft. Benning to extract environmental variables from a CART analysis for Dr. Eckhardt's tree plots. Table 14 lists the variables that we used in the analysis.

Table 14: Variables used in Loblolly Pine Decline environmental CART analysis.

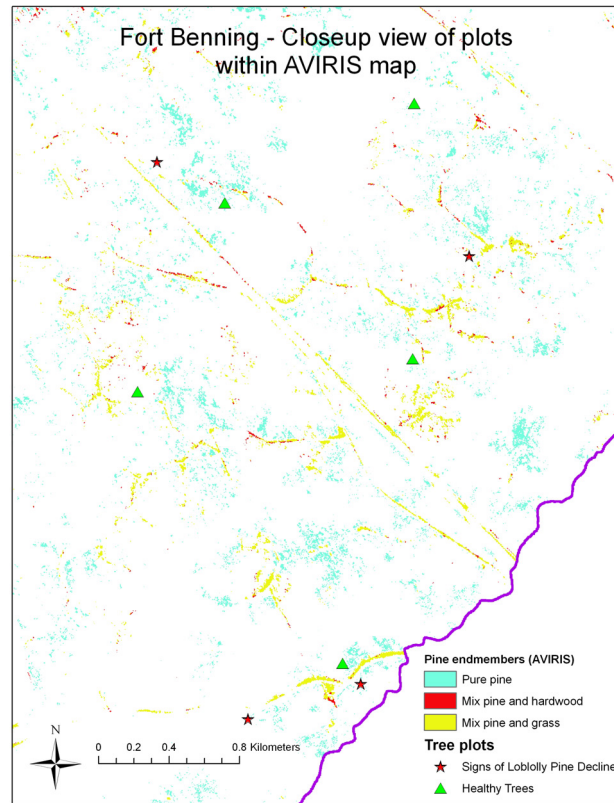| Variable	Definition	Values
elevation	Meters above sea level	meters
Aspect	Direction of greatest slope	degrees
Slope	Rise over run	percent
curvature	The degree of curvature of the terrain	0 = flat, negative = concave, positive = convex
Erosion	Does the soil survey list soils as eroded?	Yes or No
Surface	Soil surface texture	L=loam, S=sand, LS=loamy sand, SL=sandy loam, CL=clay, SCL=sandy clay loam
No40	Percent soil passing No. 40 (42mm) sieve	percent (max and min)
No200	Percent soil passing No. 200 (0.074mm) sieve	percent (max and min)
Clay	Clay content of soil	percent (max and min)
perm	The quality of the soil that enables water to move downward through the profile. Measured as the number of inches per hour that water moves downward through the saturated soil.	inches / hr
awcrank	Mean maximum value for the soil available water capacity in inches water per inches soil expressed as a volume fraction. Example: If a soil has a water capacity value of 0.20 a 10 inch zone then contains 2 inches of available water.	inch water / inch soil
minph	Mean minimum value for soil pH	0 through 14
maxph	Mean maximum value for soil pH	0 through 14
ohmrank	The amount of soluble salts in a soil. The conventional measure of soil salinity is the electrical conductivity of a saturation extract (Mohms/cm at 25C).	rank: 1 to 4

We used the Rulequest Research, SEE5 data mining program to create the CART model. The CART analysis on the GIS data was able to predict the training set within an error of 2.6% when boosting was used.

The following environmental factors were important in the CART model of LPD:

1) Aspect (the only variable of importance for Regression Trees 0 and 2): Slopes oriented in a North or Northwest direction were unfavorable for LPD as also previously described by Dr. Eckhardt. The remaining slopes seemed to be favorable for the spread of LPD. However, this may be an artifact of the small size of our sample. Of the 18 LPD plots, there were no plots on a north-facing or north-west facing slope and only one plot on a north-east facing slope. The north and northwest aspects are the only slopes "healthy" trees are found, according to the CART analysis.

2) Surface soil texture (important in Regression Trees 1, 3, and 4): The LPD was found on plots with sandy or loamy surface textures. Loamy sand and sandy clay loam had a mix of healthy and LPD trees.

3) Profile topographic curvature (important in Regression Tree 1): Topographic areas that are concave in shape are prone to LPD, again consistent with Dr. Eckhardt's (Auburn University) prior research at Ft. Benning.

4) Slope (important in Regression Trees 1, 3 and 4): Areas of steep slope ($> 6\%$) are less prone to LPD than flatter terrain areas. This result is inconsistent with Dr. Eckhardt's research, however.

The summed results for the five trees are shown in Figure 24. In the figure, nearly the entire base is identified by at least one CART regression tree. To classifying using the boosting algorithm, the outcome for each CART tree is counted as a probability and pixels with a tally of 3 or more probabilities results in CART classifying the pixels as areas prone to LPD. Figure 25 shows the final classification result using this procedure and Figure 26 shows an enlarged view.
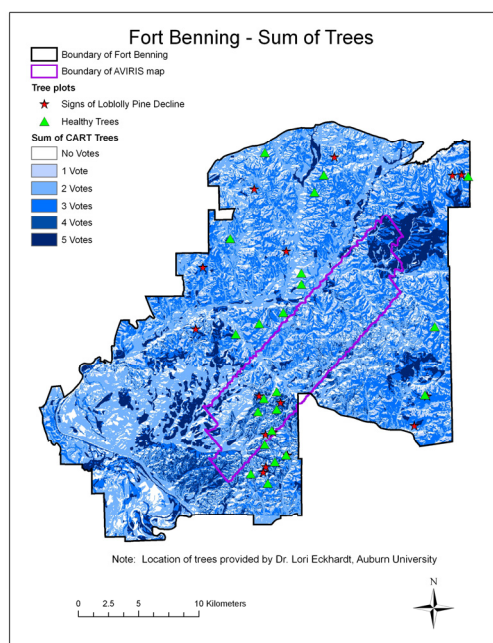
Figure 24:. Summed results for the five trees at Fort Benning
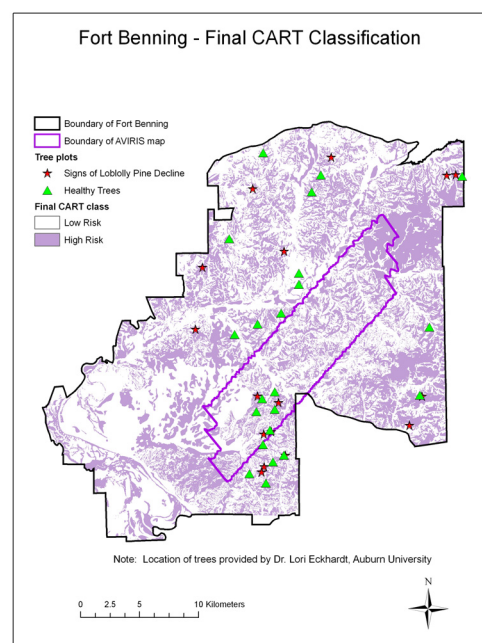
Figure 25: Final CART Classification at Fort Benning
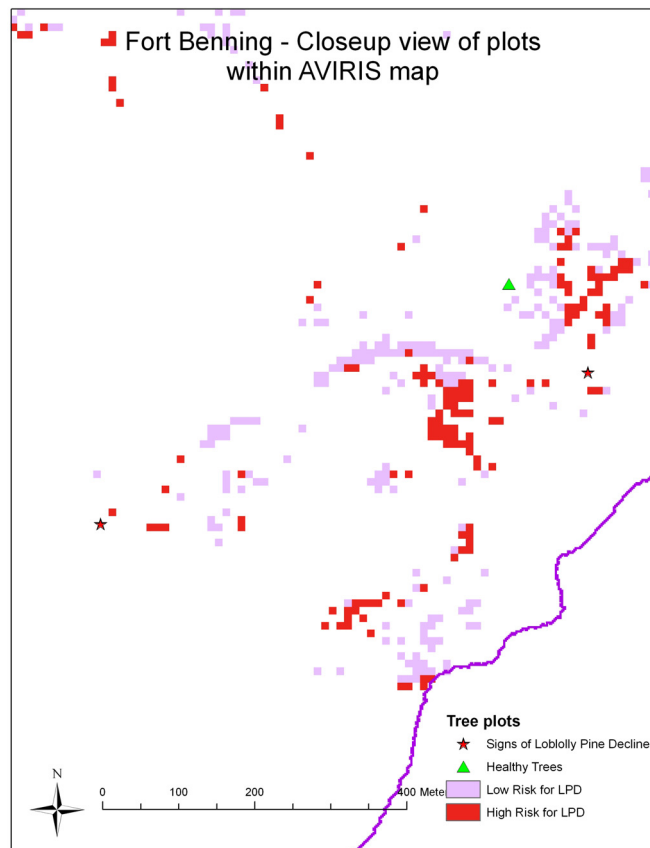
Figure 26: *Close-up view of plots within CART classification map*

5.1.5.5. Conclusions and Lessons Learned

Based on the data available, it was possible to adequately map pine species within Ft. Benning although direct mapping of tree health was poor. This is most likely due to too few LPD affected tree samples to acquire a good search relationship. When tree data were combined with GIS layers, it was found that healthy and declining pines were somewhat associated with topography and soil conditions. This combination of GIS data and hyperspectral image data holds promise for tree-level mapping of LPD.

6. Project Conclusions

To test the technology of hyperspectral remote sensing data for the detection and mapping of invasive plant species, hyperspectral imagery was acquired at six military bases, located throughout the U.S., to map twelve invasive species. Sites and species were chosen to represent a wide range of habitats and vegetation characteristics as described in Table 1. A range of classification techniques was evaluated for each species in order to identify the optimal approach. The algorithms tested were implemented in standard image processing software (Environment for Visualization of Images, ENVI) and are thus commonly available to invasive plant species management and base personnel. No other invasive plant species mapping study of this scale and diversity exists. Integrated over the six sites and twelve species, these results present an unprecedented demonstration of species-level mapping. Taken together, a consistent picture of which method performs the best under a given set of conditions emerges (Table 2).

No classifications were performed directly on the 224 band reflectance data set. Under all conditions, data reduction techniques were used, such as the minimum noise fraction (MNF), which retains the full spectrum information, and physiological indexes that target specific absorption features. This reduction step is because the wealth of spectral information provided by hyperspectral sensors, which, although necessary for species-level discrimination, overwhelms traditional classification techniques. In general, MNF transforms outperformed physiological indexes as classification inputs. This is because MNF bands, as a type of principle component analysis, integrate the entire reflectance spectrum and contain more spectral information than indexes derived from pairs of narrow bands.

Vegetation Indexes (VIs) like the normalized difference vegetation index (NDVI) perform as well as MNF when the target species is physiologically distinct from background vegetation and the chosen indexes are sensitive to that trait. For example, iceplant, an invasive succulent, was clearly distinguishable using foliar water indexes (Underwood et al., 2003). Indexes were also the most successful at mapping common reed on Carroll Island of Aberdeen Proving Ground. Only three indexes contributed to the latter classification: NDVI, the normalized difference water index (NDWI), and the photochemical reflectance index (PRI). These are metrics for vegetation cover, foliar water content, and xanthophyll cycle pigments and photosynthetic efficiency, respectively, illustrating that common reed differs from surrounding vegetation in these parameters. In a study on invasive species in Hawaii, Asner & Vitousek (2005) also found that spectrally estimated physiological parameters were sufficient to detect a structurally distinctive nitrogen fixing invasive species with unique physiological traits (i.e., elevated foliar nitrogen and water contents relative to native species). Andrew & Ustin (2006) successfully discriminated the invasive plant species *Lepidium latifolium* from co-occurring vegetation using indexes that highlighted unique reflectance in the visible wavelengths due to flowering. In short, the choice of data reduction technique will depend upon the spectral and physiological features of the target species. When the invasive plant species to be mapped possesses novel physical traits that are manifested spectrally, specific relevant spectral indexes will successfully detect the species. However, when the target species differ only subtly from background vegetation over a broader range of wavelengths, statistical dimensionality reduction tools such as the MNF transform are more appropriate.

Following data reduction, either a classification using the transformed bands from the MNF or a spectral unmixing algorithm was performed to identify target invasive plant species in the image data. Species architecture and stand characteristics determine whether a classification versus an unmixing approach is needed to define individual species. Classifications are statistical tools that define group membership of pixels. They assume that pixels are dominated by a single component and provide no information on the proportional composition of a pixel. Conversely, spectral unmixing algorithms are subpixel analyses that model each pixel as composed of fractions of several materials. Not surprisingly, classifications perform well when detecting dense monotypic vegetation while unmixing approaches are necessary for species with sparse canopies or that occur in mixed stands.

Several classification tools successfully mapped iceplant, jubata grass, kudzu, love grass, and under certain conditions, tamarisk and common reed (Table 2). In all cases, the target species occurred in large patches of robust growth relative to the pixel size. The two preferred classification algorithms were maximum likelihood (ML) and spectral angle mapper (SAM). These techniques are markedly different and their selection is probably predicated most importantly by the quantity and variability of training data. ML calculates summary statistics (mean, variance, covariance) for each pixel at all bands to generate probability distribution functions for each class. Pixels are classified to the group to which they have the highest likelihood. ML requires a large amount of training data in order to accurately estimate statistics for each class. Furthermore, classes must meet the assumptions of the ML classifier; i.e., spectra must be normally distributed. SAM requires much less training data. This algorithm compares the unclassified pixel spectrum individually to each training spectral class and calculates the angular divergence between them in spectral space, by treating the each spectrum as a large-dimensional vector. Pixels are assigned to the most similar class based on the calculated angles and assigned thresholds. Both algorithms perform well and their selection appears to be more dependent upon how much training data is available and how representative the reference spectrum is for each class rather than the spectral and physical characteristics of the species of interest.

Unmixing algorithms were used to detect fennel, cheatgrass, Russian knapweed, and, under certain conditions, tamarisk and common reed (Table 2). Each of these species presented sparse canopies and often patchy, fragmented distributions. While this problem is less important with newer high spatial resolution instruments, sparse vegetation always creates a mixed spectrum of live vegetation, plant litter and woody debris, and soil. Pixel spectra were thus mixed with litter in the case of fennel and cheatgrass, soil in the case of tamarisk, and/or other vegetation in the case of common reed. Mixtures degrade pixel spectra, reducing the spectral uniqueness of a given species relative to the background. Additionally, since plants may occur mixed with a variety of components, mixing increases the spectral variability of target species. The net result is a concurrent reduction of between-species variation and increase of within-species variation, which traditional classification approaches cannot successfully resolve.

Mixture tuned matched filtering (MTMF), the unmixing algorithm performed in all mixed cases here, is a very powerful image processing technique derived from signal processing methods. MTMFs use the MNF transformed image as input. This method has been demonstrated to be capable of detecting materials that differ only subtly from the background, such as sparse and/or

patchy distribution of invasive plant species (Parker Williams & Hunt, 2002; Mundt et al., 2005; Glenn et al., 2005). The matched filter method is constrained by a mixing model that determines the likelihood that the material is correctly classified. MTMF turned out to be as well-suited for the four sparse invasive plant species studied here as those in the references above.

The previous discussion focused on how species characteristics determine optimal mapping techniques. However, the interaction of species and site characteristics can often be just as important. For two of the species tamarisk at Yuma Proving Ground, and common reed at Aberdeen Proving Ground, different techniques were required to address site conditions. At Yuma, tamarisk occurred in two distinct zones: in dense thickets surrounding irrigated agriculture and in dry desert washes as isolated, sparse shrubs with low leaf area. When receiving irrigation outwash, tamarisk had a large and robust green canopy and much more leaf area than other desert vegetation, and was thus comparatively easy to detect. In the dry ephemeral washes, however, tamarisk was spectrally and morphologically more similar to the condition of the native desert shrubs, at least at the time of overflight. This situation was aggravated by the data being collected in late summer during the third year of a regional drought. All mapping methods successfully identified irrigated tamarisk, but MTMF was the only approach that could detect tamarisk in the dry wash habitat. Although this variation was observed spatially, similar patterns should be observable following the course of tamarisk invasion. A phreatophytic species, tamarisk's deep roots access the water table to grow more robustly than native vegetation. However, young individuals that have not yet reached deep water should resemble native species more closely than they do mature tamarisk shrubs. Advanced unmixing strategies such as MTMF are likely to be required to map the early stages of an invasion in this habitat.

At Aberdeen, common reed was mapped on two islands: Carroll and Spesutie. On Carroll Island, the ML classifier was performed on output from 19 spectral indexes (Table 3), which achieved the highest accuracies, while on Spesutie Island, MTMF was required to map this species. This requirement for different methods is corroborated by the observations of the field crews. At the time of image acquisition, common reed infestations were large, dense, and monospecific on Carroll Island but tended to be patchier and mixed with native species on Spesutie Island. Detailed knowledge of species characteristics, site characteristics, and the specific infestation characteristics at each site are considerations that should inform the choice of mapping methods. Poor classification accuracy results when using an inappropriate method but this in itself does not provide insight into why the analysis failed or what is needed to improve the accuracy of a classification.

Finally, not all species were successfully discriminated under all conditions in this SERDP study. Although targeted as management priorities, Asian mustard and hydrilla could not be mapped with the image data that was acquired. Asian mustard was senescent and or dead at the time of image acquisition due to the late summer timing of the overflight and the extreme drought conditions. It is challenging if not impossible to spectrally resolve different species of dead vegetation. Somewhat unexpectedly, dry fennel (*Foeniculum vulgare*) was successfully mapped in dry grasslands at Camp Pendleton in an autumn image using MTMF. This discrimination may be more related to reflectance differences due to canopy structural differences rather then spectral differences in the plant material. While it is possible to show that different types of dry

plant material (leaf litter, wood, bark, etc.) are spectrally different, development of the methods to differentiate these among different species has not occurred.

Hydrilla is a species of submerged aquatic vegetation, and the Aberdeen image data was not acquired with the requirements for mapping submerged species, e.g., acquisition avoiding sunglint and low tide level. Submerged aquatic invasive plant species were mapped in other studies, primarily Brazilian Waterweed plant species (*Egeria densa*) in the Sacramento delta of California (Underwood et al., 2006b; Hestir et al., 2008). See Hestir et al. (2008) for a discussion of the unique challenges and special considerations of detecting submerged aquatic vegetation. For species such as these that pose additional challenges to detection, image timing may be the determining factor. It is common knowledge that species are more easily mapped when data acquisition is timed to coincide with unique phenological stages. For Asian mustard, imagery should be acquired while the plants are living, instead of the dry season of a record drought year. In this study the project had no control of the timing of NASA's overflight schedule; however, timing of data acquisition would be of key importance for a management protocol. For hydrilla, imagery must be acquired under low tide and calm surface conditions (i.e., low wind conditions) and when sun angles are low enough to avoid sunglint in the imagery. Even when a species is not detectable at any single image date, a combination of multi-date images, by capturing phenological trajectories, may allow mapping of recalcitrant species. For example, the in-tandem analysis of both spring and summer image dates of Yakima improved mapping accuracy of cheatgrass.

Table 15: Indicators for Invasive Species Potentially Measurable by Remote Sensing

Indicator	Spatial Scale	Description	Considerations of remote sensing technology recommended with the following qualifications
Area covered by invasive species	m to km	Success highest when infestations are dense or have a uniform canopy	All full spectrum methods; physiological spectral indexes
Identification of particular invasive species	M	Possible where target species has distinct biochemical or structural characteristics compared to surrounding species, e.g., water absorption, early green-up	Requires identification of spectral bands; methods like physiological spectral indexes, continuum removal; all full spectrum methods
Trends in invasive species	m to km	Time series analysis	Multidate imagery of the same area; should be acquired at the same stage of the species life cycle and processed in the same manner
Multidate imagery to map particular invasive species	m to km	Multi-season analysis	Multidate imagery of the same area; Acquired in multiple seasons to take advantage in phenological differences between invasive plant species and native vegetation
Variability of native vegetation	M	Areas of sparse native vegetation often vulnerable to invasion. Areas of either high or low species richness are also reported to be vulnerable	Depends on the measurement scale and distribution of vegetation community; minimize within class variation and maximize between class variation (invasive plant species vs. native vegetation)
Identification of vulnerable areas based on known invasion drivers	m to km	Invasion is often greatest in riparian zones; were topographic relief is low	Overlay image data on digital terrain map
Disturbance sites (e.g., roads or natural such as landslides or post-fire)	m to km	Anthropogenic and natural disturbances create areas vulnerable to invasion	High spatial resolution or use change detection methods
Fragmentation of landscape	m to km	Edge habitats are more vulnerable to invasion than core habitats	Location of roads, construction; analyze spatial distribution and statistics (e.g., geostatistics, wavelets)
Inappropriate/excessive grazing	m to km	Intensive grazing can result in invasions	Identify areas of reduced plant cover and erosion
Proximity to sources of invasion (e.g., urban areas), military training, and transportation corridors	m to km	Urban areas provide sources of propagules (e.g., ornamental plantings) and car traffic can disperse seeds on tires	Identify man-made objects and disturbance characteristics

6.1.   Factors Relevant to Invasive Species Detection in Remote Sensing Imagery

6.1.1.   Phenological characteristics

The timing of image acquisition is an important determinant in whether an invasive species can be detected. Some invasive species become greener earlier than surrounding species in the spring or alternatively senesce earlier, e.g., cheatgrass (*Bromus tectorum*) or later, e.g., yellow starthistle (Miao et al. 2006). Again, the timing of image acquisition is important and acquiring two or more images at different times in the life cycle can improve classification results, compared to a single acquisition. Researchers mapping Chinese tallow (*Sapium sebiferum*) found that although the red senescing leaves were spectrally distinct, field observations showed that leaves turned red at different times with significant spectral variability in the leaves (Ramsey et al. 2002). Consequently they recommended three sets of data collections during the autumn to adequately map the distribution of Chinese tallow.

6.1.2.   Structural characteristics

Sometimes the architecture of the canopy, if it contrasts with the background vegetation, can be used to distinguish the target species. For example, Broom snakeweed (*Gutierrezia sarothrae*) has an erect-leaf canopy structure that makes it dark in its image response (Hunt et al. 2005b) or the contrast in canopy between invasive blue gum trees (*Eucalyptus globulus*) and the surrounding shrublands in California chaparral have proved useful for identification purposes. Figure 9 shows a vegetation map comprised of species having different growth forms.

6.1.3.   Spatial pattern of invasion

Target species whose invasion pattern has both dense and uniform canopy cover is more likely to be detectable, particularly when infestations spread over large spatial areas. For example, the rhizomatous spread of iceplant (*Carpobrotus edulis*) produces a dense, uniform cover which assists in detection (Underwood et al., 2006a) or the density of redberry juniper (*Juniperus pinchotii*) infestations are detectable compared to the surrounding rangelands (Everitt et al. 2001).

6.1.4.   Indirect measurements

The majority of studies apply remote sensing techniques to invasive species that dominate the vegetation canopy. However, many invasive species are understory species or occur in compositions of mixed species (e.g., a meadow) which present additional challenges for detection with remote sensing and requires the use of indirect methods of mapping often using other GIS data layers (Joshi et al. 2004).   The example of weeping lovegrass shown in Figure 10 was successfully mapped because it often occurred in canopy gaps and other openings at the edges of roads.

6.1.5.   Image Resolution and Image Types Needed for Identifying Invasive Species

An influential paper in the mid-1990s (Price 1994) suggested that mapping individual species from canopy reflectance may be an impossible task, given the variability within a species and the potential for species to appear similar to other ones. Today, however, the potential to use hyperspectral imagery to map species does seem feasible under some conditions and at some spatial scales. The application of remote sensing to invasive species has seen a transition from the use of imagery with a high spatial but low spectral resolution, such as panchromatic or color infrared aerial photographs, to digital images with greater spectral resolution although coarser spatial resolution, and more recently to imagery that couples both high spatial and high spectral resolution. Concurrent improvements in radiometric resolution, detector stability and calibration have also contributed to improved mapping capability.

Similar to mapping classes of vegetation for other purposes, the choice of image pixel size (spatial resolution) depends on the smallest patch that must be mapped. Researchers suggest the pixel size of the imagery should be at least one-fourth the area of the smallest patch that needs to be mapped (Hunt et al. 2005a). The advantage of finer spatial resolutions is that fewer pixels fall on the boundaries between invaded and uninvaded areas, thus increasing the number of 'pure pixels' and potentially greater mapping accuracy. The spectral resolution, or the number and width of spectral bands, provides more spectral information as the number of bands increases and thus greater likelihood of identifying the target species. The large number of published studies which have used a variety of hyperspectral image types illustrates how these capabilities relate to mapping invasive plant species.

6.1.6.   Aerial photography and videography

At the fine spatial end of the continuum of image options, aerial photographs have the benefit of being relatively inexpensive and typically have fine spatial resolution (from 0.1 m x 0.1m or 2 m x 2 m, or 0. 01 m^2 to 4 m^2). This spatial resolution meets all or most criteria for sampling even at the smallest patch size. Aerial photography (digital or film) and digital videography are particularly appropriate where the invasive species has unique visual characteristics that readily distinguish it from the surrounding vegetation, biochemical properties such as flower color, or phenological and structural characteristics like growth form (physiognomic) differences and early senescence. The use of infrared imagery can further assist in the detection of target species since there is higher reflectance in the infrared rather than the visible part of the spectrum thus allowing differences to be more pronounced. In addition, spatial variation in canopy texture (as detected by brightness variation) can contribute to species identification.
       Chinese tamarisk (*Tamarix chinensis*) has been identified with aerial photography using the unique orange-brown color prior to leaf drop, the distinctive yellow bracts of leafy spurge (*Euphorbia esula*) have also provided useful characteristics for discrimination, as well as the relatively dark green foliage of redberry juniper (*Juniperus pinchotii*) compared to surrounding species (Everitt et al. 1995; Everitt et al. 1996; Everitt et al. 2001). High resolution infrared aerial photography was used to map a number of woody invasives, including blackberry (*Rubus fruticosus*), European olive (*Olea europaea*), and pine species (*Pinus spp.*) in the Mount Lofty Ranges of South Australia and achieved an overall success rate of 74%, but noted that when the species occurred in the understory successful detection was limited to either gaps or at the fringes of the canopy (Crossman & Kochergen 2002). Visual and computer assisted

interpretation of digital infrared images was also used to successfully map invading *Acacia* species from surrounding native vegetation and other invasive species in the fynbos biome of South Africa (Stow et al. 2000). In the aquatic realm, expert interpretation of color aerial photographs were successful in mapping submerged aquatic vegetation in the Lower Chesapeake Bay, Maryland (Orth & Moore 1983). However, similar photographic methods in the Sacramento-San Joaquin delta in California were not able to provide reasonable maps of submerged aquatic invasive plant species.

While aerial photography is relatively inexpensive and can be acquired at high spatial resolution, disadvantages include extensive manual labor for processing including digitizing and georegistering of imagery as well as time-intensive interpretation requiring both skill and experience (Anderson et al. 1993; Everitt et al. 1995). Although modern digital aerial cameras preclude the need to scan photographs from film and GPS and onboard navigation systems minimize registration problems, interpretation and the need for expert interpreters remains the greatest obstacle. Furthermore, as highlighted in the studies reviewed here, its utility often relies on the visual identification of unique spectral characteristics which places significant challenges in coordinating image acquisition and field measurements at the appropriate time to capture spectral differences. Given the resolution of the imagery, along with the requirement for expert interpretation and associated manual processing time, data collection is only feasible over relatively small spatial areas.

6.1.7. Multispectral imagery

In contrast to aerial photography and videography, the use of digital multispectral imagery offers coverage over larger spatial areas, objective change detection through direct analysis of historical image archives, and the opportunity for automated image processing. Some studies have had success even using resolution as coarse as the Advanced Very High Resolution Radiometer (AVHRR) (1.1 km x 1.1 km pixels with red and near-infrared wavebands) imagery to identify invasive species. For example, identifying moderate to heavy infestations of broom snakeweed (*Gutierrezia sarothrae*) from surrounding grassland species was achieved by capitalizing on differences in phenological activity using the normalized difference vegetation index (NDVI) - a normalized ratio of red and near-infrared (Peters et al. 1992). Information on water clarity and turbidity from AVHRR images collected in different seasons were found to be accurate indicators of the locations of zebra mussels (*Dreissena polymorpha*) in Lake Huron (Budd et al. 2001). Clearly, these examples represent invasive species which have spatially extensive distributions and where background conditions remain relatively consistent over large areas.

At a medium spatial and spectral resolution, the Landsat Thematic Mapper (TM) (30 m x 30 m pixels and 6 spectral bands) imagery has been used successfully to map target invasive plant species that are spectrally or temporally unique. For example, infestations of dyers woad (*Isatis tinctoria*) were mapped using correlations with brightness, greenness, and wetness values derived from reflectance data (Dewey, Price, & Ramsey, 1991). Similarly, Systéme Pour l'Observation de la Terre (SPOT, 20 m x 20 m pixels and 4 spectral bands + 1 panchromatic band at 10 m x 10 m pixels) imagery has been used to successfully identify speargrass (*Imperata cylindrica*) invading savanna areas in Cameroon (Thenkabail 1999). SPOT has also been suggested to be an appropriate sensor for monitoring the control of emergent aquatic invasive plants such as water hyacinth (*Eichhornia crassipes*) in Bangalore, India using NDVI

(Venugopal 1998). Landsat Enhanced TM (ETM: 30m x 30 m pixels, with 6 spectral bands + 1 panchromatic band at 10 m x 10 m) has been used to locate invasive species in the understory using indirect methods. Forest canopy density and light intensity reaching the understory were derived from Landsat imagery and combined with data on grazing to map the extent of bitter bush (*Chromolaena odorata*) in lowland forests in Nepal (Joshi et al. 2006).

Multiband imagery offers some clear advantages over aerial photographs and videography but the coarse spatial and spectral resolution of both AVHRR, MODIS, and Landsat (E)TM present challenges for infestations which are neither continuously widespread, dense, nor of monospecific composition. Additionally, the 16 day repeat interval may make it difficult to target acquiring data at a specific stage in the life cycle.

6.1.8.   Hyperspectral imagery

While aerial photography and multiband remote sensing imagery have been available for many years, more recently, new technologies such as hyperspectral imagery, also known as imaging spectroscopy, allow detailed laboratory-like spectra to be acquired for each pixel in an image. Airborne and spaceborne hyperspectral instruments are characterized by many narrow contiguous spectral bands that have been used to produce high mapping accuracies and identification of individual species through the detection of specific narrow-band differences and subtle variation in reflection and absorption patterns (Aspinall et al. 2002; Vane & Goetz 1993). The large number of wavebands (e.g., 224 for Airborne Visible Infrared Imaging Spectrometer (AVIRIS) imagery) offers great potential for mapping invasive species compared to multispectral and photographic methods, especially where the target species is present at low densities. Flightlines should be oriented to cover known locations of invasive plant species and native vegetation without invasive plant species invasion. Also, different installations had some restrictions on flightline locations, which resulted in flightlines not being oriented in a consistent direction. This added complications for aligning the radiometric data to allow cross-calibration of flightlines. Flying the lines in the principal plane of the sun would reduce this analysis complication. Several commercial atmospheric calibration software programs were used. Each has some advantages and disadvantages for applications. The FLAASH program yielded poorer results in sites with high humidity compared to ACORN. The program ATCOR with its topographic module provided the best surface reflectance in mountainous terrain.

6.1.9.   Context dependence of invasive plant species mapping methods

Detecting invasive species with remotely sensed data is a challenging endeavor. All vegetation has the same basic spectral characteristics since all vegetation is composed of the same suite of biochemicals (pigments, water, cellulose, etc.). Leaf spectra are dominated by the influence of only four variables: photosynthetic pigments, water, dry matter, and a structure parameter related to leaf thickness (Jacquemoud and Baret 1990). All plant species are not spectrally unique (Price 1994) and spectral variability may be greater within than between species (Okin et al. 2001).

7. References

Anderson, G. L., J. H. Everitt, A. J. Richardson, and D. E. Escobar. (1993). Using satellite data to map false broom invasive plant species (*Ericameria austrotexana*) infestations on south Texas rangelands. Weed Technology 7:865– 871.

Andrew, M. E., and S. L. Ustin. (2006). Spectral and physiological uniqueness of perennial pepperweed (*Lepidium latifolium*). Weed Science 54:1051-1062.

Aspinall, R. J., W. A. Marcus, and J. W. Boardman. (2002). Considerations in collecting, processing, and analyzing high spatial resolution hyperspectral data for environmental investigations. Journal of Geographical Systems 4:15-29.

Asner, G. P., and P. M. Vitousek. (2005). Remote analysis of biological invasion and biogeochemical change. Proceedings of the National Academy of Sciences of the United States of America 102:4383-4386.

Berry, W.L. (1970). Characteristics of salts secreted by *Tamarix aphylla*. American Journal of Botany 57:1226-1230.

Boardman, J.W. and F.A. Kruse. (1994). Automated spectral analysis: a geologic example using AVIRIS data, north Grapevine Mountains, Nevada. In: Proceedings, Tenth Thematic Conference on Geologic Remote Sensing. Environmental Research Institute of Michigan, Ann Arbor, MI, pI-407 – I-408.

Brotherson, J.D. and V. Winkel (1986). Habitat relationships of saltcedar (*Tamarix ramosissima*) in central Utah. Great Basin Naturalist 46:535-541.

Budd, J. W., T. D. Drummer, T. F. Nalepa, and G. K. Fahnenstiel. (2001). Remote sensing of biotic effects: zebra mussels (*Dreissena polymorpha*) influence on water clarity in Saginaw Bay, Lake Huron. Limnology and Oceanography 46:213-223.

Busch, D.E. and S.D. Smith (1995). Mechanisms associated with decline of woody species in riparian ecosystems of the southwestern U.S. Ecological Monographs 65:347-370.

Card, D. H., Peterson, D.L, Matson, P.A. and J.D. Aber. (1988). Prediction of leaf chemistry by the use of visible and near infrared reflectance spectroscopy. Remote Sensing of Environment 26:123-147.

Carter, G. A. (1991). Primary and secondary effects of water content on the spectral reflectance of leaves. American Journal of Botany 78(7):916-924.

Center for the Study of Earth from Space (CSES). (1992). SIPS User's Guide, The Spectral Image Processing System, v.1.1, University of Colorado, Boulder, 74 p.

Chen, J.Y. and I.S. Reed. (1987). A detection algorithm for optical targets in clutter. IEEE Transactions on Geoscience and Remote Sensing, v. 26, no. 1, p. 65-74

Cheng, Y-B., Tom, E., and Ustin, S.L. (2007). Mapping an invasive species, kudzu (*Pueraria montana*) using hyperspectral imagery in western Georgia. Journal of Applied Remote Sensing: 1: 013514 (23 May 2007).

Crossman, N. D., and J. Kochergen. (2002). Mapping environmental invasive plant species in the Mount Lofty Ranges, South Australia, using high resolution infrared aerial photography. Pages 1-4 in H. Spafford Jacob, J. Dodd, and J. H. Moore, editors. Proceedings of the thirteenth Australian invasive plant species conference. Plant Protection Society of Western Australia, Perth, Australia.

Curran, P. J. (1989). Remote sensing of foliar chemistry. Remote Sensing of Environment 30:271-278.

DiPietro, D., S. L. Ustin, E. Underwood, 2002. Mapping the invasive plant A*rundo donax* and associated riparian vegetation using AVIRIS. Proceedings of the 11th Earth Science Airborne Workshop, Jet Propulsion Laboratory, Pasadena, CA, March 6-8, 2002.

Di Tomaso, J.M. (1998). Impact, biology, and ecology of saltcedar (*Tamarix* spp.) in the southwestern United States. Weed Technology 12:326-336.

Elvidge, C. D. (1990). Visible and near infrared reflectance characteristics of dry plant materials. International Journal of Remote Sensing of Environment 11:1775–1795.

Everitt, J.H. and C.J. Deloach (1990). Remote sensing of Chinese tamarisk (*Tamarix chinensis*) and associated vegetation. Weed Science 38:273-278.

Everitt, J.H., D.E. Escobar, M.A. Alaniz, R. Villarreal, and M.R. Davis (1992). Distinguishing brush and invasive plant species on rangelands using video remote sensing. Weed Technology 6:913-921.

Everitt, J. H., G. L. Anderson, D. E. Escobar, M. R. S. Davis, N. R., and R. J. Andrascik. (1995). Use of remote sensing for detecting and mapping leafy spurge (*Euphorbia esula*). Weed Technology 9:599-609.

Everitt, J.H., D.E. Escobar, M.A. Alaniz, M.R. Davis, and J.V. Richerson (1996). Using spatial information technologies to map Chinese tamarisk (*Tamarix chinensis*) infestations. Weed Science 44:194-201.

Everitt, J. H., C. Yang, B. J. Racher, C. M. Britton, and M. R. Davis. (2001). Remote sensing of redberry juniper in the Texas rolling plains. Journal of Range Management 54:254-259.

Fourty, T. and F. Baret (1998). On spectral estimates of fresh leaf biochemistry. International Journal of Remote Sensing 19(7):1283-1297.

Friederici, P. (1995). The alien saltcedar. American Forests 101:44-47.

Fuentes, D. A., J. A. Gamon, H.-L. Qiu, D. A. Sims, and D. A. Roberts. (2001). Mapping Canadian boreal forest vegetation using pigment and water absorption features derived from the AVIRIS sensor. J. Geophys. Res. 106:33565-33577.

Gao, B.-C. (1996). NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sensing of Environment 58:257-266.

Glenn, N. F., J. T. Mundt, K. T. Weber, T. S. Prather, L. W. Lass, and T. S. Pettingill. (2005). Hyperspectral data processing for repeat detection of small infestations of leafy spruge. Remote Sensing of Environment 95:399-412.

Green, A. A., Berman, M., Switzer, P. and Craig, M. D. (1988). A transformation for ordering multispectral data in terms of image quality with implications for noise removal. IEEE Transactions on Geoscience and Remote Sensing, 26(1):65-74.

Harsanyi, J.C. and C.I. Chang. (1994). Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach. IEEE Transactions on Geoscience and Remote Sensing, 32:779-785

Hestir, E.L., S. Khanna, M.E. Andrew, M.J. Santos, J.H. Viers, J.A. Greenberg, S.S. Rajapakse, S.L. Ustin. (2008). Identification of invasive vegetation using hyperspectral remote sensing in the California Delta ecosystem. Remote Sensing of Environment (in press).

Hunt, R., R. Hamilton, and J. Everitt. (2005a). A weed manager's guide to remote sensing and GIS (rsac-0077). Mapping weed infestations using remote sensing. http://www.fs.fed.us/eng/rsac/invasivespecies/documents/mapping.pdf Pages 1-8.

Hunt, R., R. Hamilton, and J. Everitt. (2005b). A weed manager's guide to remote sensing and GIS (rsac-0077). What invasive plant species can be remotely sensed? http://www.fs.fed.us/eng/rsac/invasivespecies/documents/Distinguishing_characteristics3.pdf. Pages 1-5.

Jacquemoud, S., and F. Baret. (1990). PROSPECT - A model of leaf optical-properties spectra. Remote Sensing of Environment 34:75-91.

Jackson, R. D. and A. R. Huete (1991). Interpreting vegetation indices. Preventive Veterinary Medicine 11:185-200.

Johnson, S. (1987). Can tamarisk be controlled? Fremontia 15:19-20.

Joshi, C., J. de Leeuw, J. van Andel, A. K. Skidmore, H. D. Lekhak, I. C. van Duren, and N. Norbu. (2006). Indirect remote sensing of a cryptic forest understorey invasive species. Forest Ecology and Management 225:245-256.

Joshi, C., J. de Leeuw, and I. C. van Duren. (2004). Remote sensing and GIS applications for mapping and spatial modelling of invasive species. Pages 669-677. Proceedings for ISPRS, Istanbul, Turkey.

Kruse, F.A., Lefkoff, A.B., Boardman, J.W., Heidebrecht, K.B., Shapiro, A.T., Barloon, J.P., and A.F.H. Goetz. (1993). The spectral image processing system (SIPS) – Interactive visualization and analysis of imaging spectrometer data. Remote Sensing of the Environment 44:145-163

Lichtenthaler, H. K., M. Lang, M. Sowinska, F. Heisel, and J. A. Miehe. (1996). Detection of vegetation stress via a new high resolution fluorescence imaging system. Journal of Plant Physiology 148:599-612.

Loope, L.L., P.G. Sanchez, P.W. Tarr, W.L. Loope, and R.L. Anderson. (1988). Biological invasions of arid land nature reserves. Biological Conservation 44:95-118.

Mal, T.K. and L. Narine (2004). The biology of Canadian invasive plant species. 129. Phragmites australis (Cav.) Trin. Ex Steud. Canadian Journal of Plant Science 84:365-396.

Miao, X., P. Gong, S. Swope, R. L. Pu, R. Carruthers, G. L. Anderson, J. S. Heaton, and C. R. Tracy. (2006). Estimation of yellow starthistle abundance through CASI-2 hyperspectral imagery using linear spectral mixture methods. Remote Sensing of Environment 101:329-341.

Morisette, J.T. Jarnevich, C.S., Ullah, A., Cai, W., Pedelty, J.A., Gentle, J.E., Stohlgren, T.J., and Schnase, J.L. (2006) A tamarisk habitat suitability map for the continental United States. Frontiers in Ecology and the Environment 4: 11-17.

Mundt, J.T., Glenn, N.F., Weber, K.T., Prather, T.S., Lass, L.W., & Pettingill, J. (2005). Discrimination of hoary cress and determination of its detection limits via hyperspectral image processing and accuracy assessment techniques. Remote Sensing of Environment 96:509-517

Nagler, P. L., C. S. T. Daughtry, and S. N. Goward. (2000). Plant litter and soil reflectance. Remote Sensing of Environment 71:207-215.

Okin, G.S., D.A. Roberts, B. Murray, and W.J. Okin. (2001). Practical limits on hyperspectral vegetation discrimination in arid and semiarid environments. Remote Sensing of Environment 77:212-225.

O'Neill, M., S. L. Ustin, S. Hager, and R. Root. (2000). Mapping the distribution of leafy spurge at Theodore Roosevelt National Park using AVIRIS. Pages 339-348 *in* Proceedings of the 9th AVIRIS Earth Science Workshop,. JPL Publication 00-18. Pasadena, CA: NASA Jet Propulsion Laboratory.

Orth, R. J., and K. A. Moore. (1983). Submerged vascular plants: techniques for analyzing their distribution and abundance. Marine Technology Society Journal 17:38-52.

Parker Williams, A., and E. R. Hunt. (2002). Estimation of leafy spurge cover from hyperspectral imagery using mixture tuned matched filtering. Remote Sensing of Environment 82:446-456.

Peñuelas, J., F. Baret, and I. Filella. (1995). Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance. Photosynthetica 31:221-230.

Peñuelas, J., J. Pinol, R. Ogaya, and I. Filella. (1997). Estimation of plant water concentration by the reflectance Water Index WI (R900/R970). International Journal of Remote Sensing 18:2869-2875.

Peters, A. J., B. C. Reed, M. D. Eve, and K. C. McDaniel. (1992). Remote sensing of broom snakeweed (*Gutierrezia sarothrae*) with NOAA-10 spectral image processing. Weed Technology 6:1015–1020.

Pontius, J., R. Hallett, and M. Martin. (2005). Using AVIRIS to assess hemlock abundance and early decline in Catskills, New York. Remote Sensing of Environment 97:163-173.

Price, J. C. (1994). How unique are spectral signatures? Remote Sensing of Environment 49:181-186.

Rahman, A. F., J. A. Gamon, D. A. Fuentes, D. A. Roberts, and D. Prentiss. (2001). Modeling spatially distributed ecosystem flux of boreal forest using hyperspectral indices from AVIRIS imagery. J. Geophys. Res. 106:33579-33591.

Ramsey, E. W., G. A. Nelson, S. K. Sapkota, E. B. Seeger, and K. D. Martella. (2002). Mapping Chinese Tallow with color-infrared photography. Photogrammetric Engineering & Remote Sensing 68:251-255.

Research Systems. (2002). ENVI Tutorials. December 2002 Edition. Research Systems, Inc. A Kodak Company.

Sanderson, E. W., M. Zhang, S. L. Ustin, and E. Rejmankova. (1998). Geostatistical scaling of canopy water content in a California salt marsh. Landscape Ecology 13:79–92.

Serrano, L., J. Peñuelas, and S. L. Ustin. (2002). Remote sensing of nitrogen and lignin in Mediterranean vegetation from AVIRIS data: decomposing biochemical from structural signals. Remote Sensing of Environment 81:355-364.

Smith, S.D, D.A. Devitt, A. Sala, J.R. Cleverly, and D.E. Busch. (1998). Water relations of riparian plants from warm desert regions. Wetlands 18:687-696.

Stocker, A.D., Reed, I.S. and X.Yu. (1990). Multidimensional signal processing for electrooptical target detection. Proceedings, SPIE International Society of Optical Engineers. Vol. 1305.

Stow, D., A. Hope, D. Richardson, D. Chen, C. Garrison, and D. Service. (2000). Potential of colour-infrared digital camera imagery for inventory and mapping of alien plant invasions in South African shrublands. International Journal of Remote Sensing 21:2965 - 2970.

Underwood, E., S. Ustin, and D. DiPietro. (2003). Mapping nonnative plants using hyperspectral imagery. Remote Sensing of Environment 86:150-161.

Underwood, E. C., S. L. Ustin, and C. M. Ramirez. (2006a). A comparison of spatial and spectral image resolution for mapping invasive plants in coastal California. Journal of Environmental Management 39:63-83.

Underwood, E., M.J. Mulitsch, J.A. Greenberg, M. L. Whiting, S.L. Ustin and S.C. Kefauver. (2006b). Large Scale Mapping of Invasive Aquatic Vegetation in the Sacramento-San Joaquin Delta Using Hyperspectral Imagery. Ecological Monitoring and Assessment 121 (1-3): 47-64.

Underwood, E.C. and S.L. Ustin. (2007). Trends in invasive species, Chapter 11, Sourcebook on Remote Sensing and Biodiversity Indicators, UN Secretariat of the Convention on Biological Diversity (CBD), In press.

Ustin, S. L. and S. Jacquemoud. (2002). Leaf optical properties. Digital Photobiology Compendium. www.photobiology.info

Vane, G., and A. F. H. Goetz. (1993). Terrestrial imaging spectrometry: current status future trends. Remote Sensing of Environment 44:117–126.

Vitousek, P.M. (1986). Biological invasions and ecosystem properties: can species make a difference? pages 163-176 in H.A. Mooney and J.A. Drake, editors, Ecology of Biological Invasions of North America and Hawaii. Springer, NY.

Vitousek, P.M. (1990). Biological invasions and ecosystem processes: towards an integration of population biology and ecosystem studies. Oikos 57:7-13.

Yu, X., Reed, I.S., and A.D. Stocker. (1993). Comparative performance analysis of adaptive multispectral detectors. IEEE Transactions on Signal Processing 41(8).

Zarco-Tejada, P. J. (1998). Optical indices as bioindicators of forest sustainability. Research Evaluation Course. York University, Toronto.

Zavaleta, E. (2000a). The economic value of controlling an invasive shrub. Ambio 29: 462-467.

Zavaleta, E. (2000b). Valuing ecosystem services lost to Tamarix invasion in the United States. pages 261-300 in H.A. Mooney and R.J. Hobbs, editors, Invasive Species in a Changing World. Island Press, Washington DC.

Appendix A: Project Publications

Cheng, Y-B., E. Tom, and S.L. Ustin. (2007). Mapping an invasive species, kudzu (*Pueraria montana*), using hyperspectral imagery in western Georgia. Journal of Applied Remote Sensing 1: 013514 (23 May 2007).

DiPietro, D., S. L. Ustin, E. Underwood. (2002). Mapping the invasive plant *Arundo donax* and associated riparian vegetation using AVIRIS. Proceedings of the 11th Earth Science Airborne Workshop, Jet Propulsion Laboratory, Pasadena, CA, March 6-8, 2002.

Noujdina, N. and S.L. Ustin (2007). Mapping cheatgrass (*Bromus tectorum*) infestation areas using multi-date AVIRIS data. Weed Science (in press)

Olmstead, K., S. Ustin, and T. Ernst (2004). Using CART modeling and AVIRIS imagery for assessing risk of weed invasion at Vandenberg air Force base. Proceedings of the 13th Earth Science Airborne Workshop, Jet Propulsion Laboratory, Pasadena, CA 28 March-April 1, 2004.

Underwood, E., S.L. Ustin, and D. DiPietro. (2002). Mapping non-native plants using hyperspectral imagery. Proceedings of the 11th Earth Science Airborne Workshop, Jet Propulsion Laboratory, Pasadena, CA, March 6-8, 2002.

Underwood, E. and S.L. Ustin. (2003). Mapping non-native species using hyperspectral imagery. 2003. Remote Sensing of Environment 86: 150-161.

Underwood, E. C., S. L. Ustin, and C. M. Ramirez. (2006). A comparison of spatial and spectral image resolution for mapping invasive plants in coastal California. Journal of Environmental Management 39:63-83.

Underwood, E. and S. Ustin. 2007. Trends in Invasive Alien Species. Pages 161-177 in Strand, H., Höft, R., Strittholt, J., Miles, L., Horning, N., Fosnight, E., editors. Sourcebook on Remote Sensing and Biodiversity Indicators. UN Secretariat of the Convention on Biological Diversity, Montreal, Technical Series no. 32, 201 pages.

Ustin, S.L., D. DiPietro, K. Olmstead, E. Underwood, and G.J. Scheer. (2002). Hyperspectral Remote Sensing for Invasive Species Detection and Mapping. Proceedings of International Geoscience and Remote Sensing Symposium, IGARSS-2002. Toronto, Canada. June 24-28, 2002.

Appendix B: Mapping Invasive Plant Species Using Hyperspectral Imagery
A Hyperspectral User's Guide

By

Susan Ustin
University of California Davis
2 January 2008

This guide is designed to provide a quick summary of the steps involved in planning a hyperspectral image data collection and analysis of the data to produce maps of invasive plant species. It is often impossible to meet all criteria and there are numerous tradeoffs that must be considered and balanced to obtain the best compromise for the purpose.

1. Mission Planning

1.1.1. Identify what structural and/or biochemical differences are likely to be present between the target invasive plant species and the native vegetation it is invading

- Differences in growth form (grass vs. shrub, shrub vs. tree, conifer vs. broadleaf tree, etc.)
- Differences in plant height (e.g., giant cane vs. annual grasses or low stature shrubs)
- Differences in plant density (uniform, patchy, partial cover vs. total cover)
- Differences in leaf area density (sparse leaves vs. dense leaves)
- Differences in composition (i.e., how heterogeneous the distribution of species is in the habitat: uniform to clumped)

Structural differences enhance the ability to detect invasive species. It will be nearly impossible to accurately map species without some differences in their structure.

1.1.2. Conduct a reconnaissance of the proposed flight lines to determine whether the area includes the range of vegetation types and site conditions being invaded and examples that are several pixels in extent, of both "pure" native species and "pure" invasive species and areas where they are mixed.

1.1.3. How diverse is the installation? Does it cover a wide elevation range and/or a mosaic of ecosystem types?

If the installation covers several distinct ecosystem types, it is best to stratify the data into these types at the first stage of analysis and treat them separately. The final result can be merged back together into one map encompassing the entire installation.

Is a pre-existing vegetation map of the installation available that is considered to be reasonably good? This can be used with the elevation map to create the different strata for the analysis. If

not, or if the map is not current, consider doing a simple land cover classification using a Landsat Thematic Mapper image. This data is available from USGS and is inexpensive if data older than 2 years is used.

1.2.1. Identify timing of image acquisition.

Data should be acquired near midday unless the target invasive species is aquatic and then because of sunglint, data is best acquired earlier or later in the day (around 9-10am or 3-4pm). Ideally, the sky should be completely free of clouds, including high cirrus clouds, as they degrade the spectral quality by attenuating the signal. The best quality data is acquired near the summer solstice when the sun angle is highest in the sky, since shadows are minimized.

In some cases, structural differences are most apparent when more shadows are present and the observation goal is to maximize shadows, e.g., when plant spacing is varied or forests of contrasting even-aged vs. multiple age structures.

1.2.2. Identify any likely differences in phenological timing between the native and invasive species. Are there differences in timing of leaf on/leaf off, flowering, or fruiting that could be used to identify the species?

Data acquired in multiple seasons will increase the accuracy of the map. If there is a time when the native species are growing and the invasive one is dormant that will improve their separation. The best time to acquire data is when the invasive species is still growing and the native species is dormant or partially dormant. These periods might be near the beginning or/and end of the full growing season.

It is very difficult to schedule an overflight to meet the timing of a very brief biological event. Also, it is unlikely that all individuals of the species to be identified in an area would be at the same phenological stage at the same time. Therefore slowly changing differences that might be present for a month or more are the preferred timing in order to get most of the species present in that growth stage..

1.3.1. Will the entire installation be flown with hyperspectral imagery?

If so, fly in N-S orientation or in the solar principal plane (in line with the sun). This minimizes shadows and differences in solar brightness across flightlines and between adjacent flightlines, making the data easier to analyze. To avoid gaps between flightlines, plan on 20-25% overlap between adjacent flightlines.

1.3.2. Only part of the base will be flown. Can flightlines be planned that cross both invaded and non-invaded native vegetation?

If practical, fly in N-S orientation or in solar principal plane.

1.4. Identify the pixel size needed to map the invasive plant species. What is a typical patch size of the invasive species and of the native species? Is the native vegetation reasonably homogeneous, that is, are the mix of species about the same over its distribution or are species clumped in different parts of the territory, making it inhomogeneous?

1.4.1. monoculture, several $>400m^2$ in size
1.4.2. monoculture, $100m^2$ in size
1.4.3. mixed species, patches never more than $50m^2$ in size

1.4.4. Consider acquiring some preliminary data to develop a krigging model of spatial dependence to use to determine the optimal size. If the pixel size is much smaller than the heterogeneity of the native community, there will be difficulty in developing a contrast between the invasive species and the invaded community. If the pixel size is much larger than the patch size of the invasive species then it will be hard to spectrally characterize. Sub-pixel methods work best if the native and invasive species are spectrally distinct, therefore pixel sizes are best if smaller than the typical invasive species patch size. Typically pixel sizes of 3m x 3m or 4m x 4m were adequate for most types of species.

1.4.5. If the native community is heterogeneous, consider treating it as multiple background types, rather than a homogeneous type in comparison to the invasive species.

1.4.6. Data volume increases exponentially as pixel size is reduced. The signal to noise is poorer (due to fewer photons detected) with small pixels so the spectral differences between the invasive species and the native species must be greater to maintain the same level of accuracy of detection. Pixel to pixel heterogeneity is greater (e.g., shadows vs. sunlight ground and foliage) as pixel size decreases. Therefore, smaller pixels will not necessarily improve the analysis and goal should be to collect largest pixel size that is likely to detect "pure" pixels of the invasive species and average the variability of the native plant community.

1.4.7. Most hyperspectral imagers are of the line scanner technology. This results in significant potential spatial distortion in the airborne data (from roll, pitch, and yaw) that must be corrected to actual geographic position to be able to use ground data or other GIS data with the analysis. If pixel size is too small it becomes difficult to spatially register the data to geographic coordinates. How small depends on several factors, including how uniform the vegetation is and whether there are recognizable "tie points" in the data (e.g., road intersections, buildings, etc.). For example, it is much easier to identify geographic tie points on a base with many recognizable features and roads than a large undisturbed natural landscape, e.g., you might find in Alaska. For the U.S., $1m^2$ pixel orthophotos are available for registering the imagery.

1.5. Collect GPS data locating the invasive species and pure patches of the native species.

1.5.1. For most plant communities, unless there is a major disturbance, once a plant community has become invaded, the invasive species remains in place. The invasive species may continue to spread over time, so it will continue to invade new areas, but will not change rapidly within a growing season. Therefore, GPS data locating invasive and native plant species can be acquired over a period of time, although recognizing that the data can become incorrect, if too much time

has passed. Nonetheless, this lets base personnel make observations of plant locations over an extended period that can be used in the analysis for training and validation data.

1.5.2. Collect GPS point data by standing in the center of a "patch" of the invasive species or native vegetation and record additional information with the point, including the estimated size of the patch or dimensions in cardinal directions, how uniform the vegetation in the patch is, it's percent cover (in 10% increments), and if it is a mix, how much of each species is present. Also record what stage in the life cycle is it at (vegetative, flowering, fruiting, senescent, dead, etc.).

1.5.3. If it is a native vegetation patch, record species presence and if it is not more or less evenly mixed, what species is dominant and collect other data on cover and life stage.

1.6. Acquiring digital photos at all locations is useful in case there are questions later during the image analysis. Standardize how the data is collected to make it easier to interpret later, e.g., from the GPS center point with the camera direction to the north at 1.5m height. If the vegetation is low, a downward looking photo at 1.5 or 2m height is also useful. These data can be linked to the GPS locations in the GIS for the project.

2. Field Data Collection During the Overflight

2.1. Record GPS locations of invasive species and native vegetation at as many sites as possible. It is not necessary that they are concurrent with the overflight, so this data can be acquired over an extended period.

2.2. During the overflight or in days immediately preceding or following, if a field spectrometer is available, obtain near solar noon (between 10am - 2pm standard time) above canopy measurements of the dominant species within the flightlines. The spectrometer foreoptics should be leveled so that the data is acquired directly above the canopy. Collect 5-10 spectra per area and record species, plant condition and GPS location.

2.3. Collect as many spectra from as many as possible locations along the flightlines within project limitations. This will help the analysis if the conditions vary within the flightlines.

2.4. Collect spectra of invariant calibration targets. These should be parking lots, sand, bare dry soil, or rock outcrops. These can be used with the atmospheric correction programs to improve the calibration. Programs don't do a good job with dust and aerosols and have problems when a wide range in elevation is present. This step makes the image data more suitable for using field measured spectra in the spectral library. These field data also provide independent measurements to evaluate how well calibrated to reflectance the images are.

3. Preliminary Image Data Analysis

3.1. Use an atmospheric correction algorithm to calibrate the data to apparent surface reflectance.

3.1.1. There are several commercial packages available for correcting the data to reflectance; examples are ATCOR, ACORN, and FLAASH. All programs produce artifacts and unless you are comparing data from various sites or want to use a spectral library to identify species, these errors probably will not affect the results.

3.1.2. To make a mosaic of several flightlines, the data will need to be cross-calibrated or the different flightlines in the mosaic will have brightness differences that will be observable. This is done by selecting one line as the "reference" and calibrating the other lines to it. This is done by developing a series of empirical linear regressions for each flightline relative to the reference image for each spectral band. Pixels located in the overlapping regions, where the reflectance of the reference and other image, can be used to develop band-by-band regressions. These regressions are then applied to the fitted images to bring them into spectral alignment with the reference image.

3.1.3. If there is extreme topographic variation in the imagery or climate conditions, the data sets could be stratified (segmented) into different regions and different calibrations applied. Such a case might be present if you had tall mountains next to the ocean, with a rain shadow on the interior side.

3.1.4.. If a field spectrometer is available, the accuracy of the atmospheric correction can be improved if "invariant" calibrations sites can be located in the imagery. It is necessary to locate at least two sites, a darker and a lighter material and to make spectral measurements of their surfaces along with GPS location of the measurements. An example of the type of surfaces would be a cement/asphalt parking lot or the flat roof of a building. These materials would be located in the imagery and a regression between the "true" field reflectance and the image "pixel" reflectance would be constructed. This band-by-band regression would be applied to all pixels in the image. For example in the program ACORN, they term this step "effort" smoothing. Another option is to measure a target continually with the field spectrometer during the overflight, and recording the time of each measurement. This way a pixel to pixel calibration could be performed in cases where there was a significant change in the atmosphere, for example with a weather front moving in or between two different flight acquisitions, if all the data could not be collected at one time.

3.2. Use a base GIS data layer to register the hyperspectral imagery to geographic coordinates.

3.2.1. Registration might be improved or/and the accuracy validated by measuring the GPS locations at recognizable places in the imagery.

3.3. Using GIS layers, eliminate all materials in the images that are not relevant to the map of the invasive species. This reduces the processing time and improve the accuracy of the final map.

3.3.1. Mask water bodies, impervious surfaces, buildings, and other man-made objects.

3.3.2. Determine if any vegetation types can be eliminated because the invasive species is not going to occur in that type, e.g., agriculture or forest? Use pre-existing vegetation map or the land cover map created for this project.

3.4. Using the imagery available, eliminate before hand any land cover types that are not relevant to the analysis.

3.4.1. Use a vegetation index like NDVI or EVI to find a threshold that eliminates non-vegetated areas if the invasive species is only present in vegetated pixels or eliminates densely vegetated pixels if the invasive species is found only at low density. Typically an NDVI of <0.2 will locate non-vegetated pixels and > 0.6 will locate high density vegetation.

3.4.2. Perform a preliminary classification according to growth form type: grass, shrubs, conifers, broadleaf trees; and use it to eliminate any classes that are not potentially composed of the target native plant community type and/or invasive species.

3.4.3. If the invasive species is a riparian zone invader, using a digital elevation model to locate streams and then selecting a buffer mask of the appropriate size, e.g., 100m on each side of the stream can be used. Likewise, an elevation limitation or a soil type limitation can be used to eliminate locations where the invasive species would not be present.

4. Hyperspectral Image Analysis

The ENVI software has a method they term the "Spectral Hourglass Wizard" which defines some basic steps in image analysis that are useful for locating and mapping invasive plant species. Use the Spectral Hourglass Wizard to guide you step-by-step through processing flow to find and map image spectral endmembers. The wizard displays detailed instructions and useful information for each function. The hourglass processing flow uses the spectrally over-determined nature of hyperspectral data to find the most spectrally pure, or spectrally unique, pixels (i.e., endmembers) within the data set, to map their locations and sub-pixel abundances.

4.1. Data Compression Steps:

Minimum Noise Transform

The MNF transform is used to determine the inherent dimensionality of image data, to segregate and equalize the noise in the data, and to reduce the computational requirements for subsequent processing. Typically the first transformed band or sometimes the first and second will relate to albedo or brightness differences and have a spatial pattern that grades across the image or from end to end and not be much affected by variability on the ground. These can be discarded from a classifier, despite containing most of the image variance. There will generally be from 10-30 bands out of the original data set of hundreds of bands, which have coherent spatial patterns, i.e., not just speckle or noise patterns, containing the most significant information of the imagery.

4.2. Application of vegetation indexes

A large number of spectral vegetation indexes (VIs) have been developed that capture differences in pigment composition, water content and dry matter in the plant canopy. If invasive species have biochemical differences relative to background vegetation (e.g., higher nitrogen, pigment, or water content, or more dead/dry matter in the canopy). This study used 19 indexes that were found to contain information over a wide range of data sets.

4.3. These MNF bands alone or VIs alone, or combined can be input into a supervised classification using training data to classify the imagery into cover types, where one or more types are the invasive species of interest.

MNF bands with training data are effective at producing maps of many invasive species. However, the MNF transforms are dependent on the variance of the specific data set and are not transferable to another site or to another data collection of the same location. However, the location of the training sites could be re-used in another dataset to compare the maps for changes in the distribution of the invasive species over time.

4.4. Locating pixels of invasive species.

4.4.1. Field Located GPS points.
GPS locations can be used to identify pixels in the image of native and invasive species. The pixel spectra can be added into a spectral library to be used to identify the species using spectral matching techniques.

4.4.2. Pixel Purity Index
The PPI is a method to find pure pixels from the image. The PPI is computed by repeatedly projecting n-D scatter plots on a random unit vector and a threshold value is used to define how many pixels are marked as extreme at the ends of the projected vector. The threshold value should be approximately two to three times the noise level in the data, which is 1 when using MNF transformed data. By understanding this small collection of the purest pixels, it is possible to understand the factors that control the variance in the imagery and whether the invasive species is distinctive. The PPI maps locations that should be visited for ground truth validation and spectral measurements in the field.

4.5. n-D Visualizer.

Locate pixels with GPS to select spectra in order to create a spectral library of known materials from the image. Compare them to the PPI pixels to determine if any of them are spectrally unique. Display known pixel spectra in n-D Visualizer to see if there are spectral projections where they are distinct.

If there are NO projections where the invasive species is distinct, it will not be possible to differentiate them unless it is possible to eliminate additional parts of the data set through more stratification. If this removes pixels that are spectrally similar to the invasive species, it will improve the detection. However, there is also the possibility that the invasive species is also being removed with the mask and tests must be done to determine if this is a problem. It is still

possible to use this segmented method and sequentially map the invasive species using these more restricted ranges of variation and at the end of these steps, create a mosaic image of the invasive species over the entire area.

4.6. Classifications using Mixture Tuned Matched Filtering (MTMF), and Spectral Unmixing, and Spectral Angle Mapper (SAM) methods as described in the final report or in the ENVI software help program.

There is no single method that is best for all conditions. The Spectral Unmixing methods provide a subpixel cover estimate of the invasive species for each pixel. If the invasive species is located nearly everywhere, this is a good choice of method to use. This method requires that all spectral types are known and also part of the mixing analysis. If the plant spectra are too similar, the method does not work well. Also, if the invasive species is not present in many pixels, i.e., many pixels have 0% cover of the invasive species, it is likely to be overestimated. A more recent method, available by ftp from Dr. Dar Roberts at University of California Santa Barbara (termed "multiple endmember spectral mixture analysis, or MESMA), addresses this limitation. However, use of this method requires significant investment in obtaining spectra of all types present in the image in contrast to standard mixing that uses a one or two "average" types to represent all vegetation in the image.

The SAM provides maps of how closely pixels match a reference spectrum. If there is little variance in the spectra between pixels this method works well. But if there are a range of conditions, e.g., might be found if soil moisture varied in different parts of the image, then this method runs into problems. Additionally, the method produces a continuous range of similarity from exactly matching the reference to completely different. The analyst must specify a threshold cutoff where the similarity value is too low to be considered a good match. This is typically done by comparison of the SAM values to the training and validation data and selecting a "best fit" cutoff.

The Mixture Tuned Matched Filter (MTMF) which uses the MNF transformed bands as inputs provided the most consistently correct maps of the invasive species, even under the most challenging conditions. This method produces two images a Matched Filter image, which determines how similar the pixel is to the reference spectrum and an "infeasibility" image that estimates the likelihood that the classification is correct. Where the MF is high and the Infeasibility is low provides a good fit to the data. The analyst must set these thresholds and again, it is done using the field training and validation data to determine the optimal threshold. This is the method that has been used most frequently in the published literature at this time.

4.7. Other methods

We originally proposed to use a newer statistical analysis procedure termed "support vector learning machines" (SVLM) to map invasive species. It was not used in this study as the commercial programs were sufficient in most cases to obtain good results and because the SVLM methods were not practical outside of a few laboratories testing these methods. This is still the case for this method although it has been tried in several studies and does appear to be a promising method for difficult mapping cases. Because the instrument capabilities and the

spatial resolution have also improved over this period of time, resulting in large improvements in mapping using established methods, the SVLM method remains relatively under-explored at this time. Nonetheless, if it was incorporated into standard image processing packages, it could prove to be a valuable tool for detecting invasive species, particularly where differences between the invading species and the background species are minimal.

We used Classification and Regression Tree (CART) models to identify pixels at risk for invasion by iceplant based on environmental conditions (e.g., soil type, topography, distance from water, etc.) or human-caused conditions (roads, fires, buildings, etc.). This type of modeling can extend the analysis beyond the immediate image data (assuming that the relevant GIS layers are present) and can provide a basis for predicting direction and location of potential spread.

5. Validation of the Invasive Species Map

5.1. Assuming that some field GPS locations were not used to develop the maps, the remaining data can be used to start the validation process by determining the accuracy of the map classes. This step is usually iterated with the analysis to get the best final product.

5.1.1. This is typically done using a matrix with the field classes on one axis and the image pixel classes on the other. This will tell you how often the classification is correct, and where image predicted classes are wrong, including which classes were confused with the correct class. This information is very useful to help improve the classification overall by understanding where mistakes occur.

5.1.2. Because the location of the pixels in the matrix are also known, it is possible to determine where in the image the classification was wrong. For example, if errors are located in one part of the flightline, it may be possible to deduce what is wrong (a missing endmember, atmospheric calibration problem, etc.).

5.2. Additional Field Data Collection for Validation

5.2.1. At the stage where the image map appears to be reasonable, the next step is to take the map to the field to acquire additional independent data on the classification accuracy.

5.2.2. Because of the large number of pixels in an image (millions), it is not feasible to do a completely random sampling design and test the entire area of the flightline(s). Alternate methods use combinations of approaches.

5.2.3. Using a hand-held GPS/computer with the image map displayed, locate boundaries on the image and go to these areas and either walk the boundary to determine if it is correctly mapped or walk transects that cross the boundary and record species and plant density.

5.2.4. Locate unusual spots or areas where you think the classification is mistaken to determine what is at these locations. An example might be a patch forming a geometric shape like a circle

or rectangle. Such areas could be disturbed habitats now invaded by the target species or some other anomaly that is mistaken for the invasive species.

5.2.5. Use the species map and stratifications and go to as many locations as possible that are scattered over the full extent of the flightline(s) that are identified for the invasive species to verify the classification.

5.2.6. Vegetation maps are dynamic. Once the map is accepted as the base, it is still important to keep validating it and updating it as new information becomes available and as the invasive species continues to spread into new areas.

6. Post-analysis Review and Assessment

6.1. Conduct a formal review of the invasive species mapping process to evaluate how methods can be improved in future activities.

6.1.1. If the invasive species was not successfully mapped, it is important to attempt to determine what was the basis of the failure to avoid similar problems in the future.

6.1.2. If the results were successful, what steps can be improved or made more efficient?

6.1.3. Use the maps to develop a model to predict how management will affect the spread of the invasive species or to predict where the invasive species will spread.

Appendix C: SERDP Tutorial

SERDP Hyperspectral Tutorial:
Mapping Invasive Plant Species with Remote Sensing Data

California Space Institute
Center of Excellence
University of California Davis
http://calspace.ucdavis.edu/

Table of Contents

1 Introduction

1.1 Format

The tutorial format is a combination of slide presentations and hands-on instruction. A training manual, computer workstation and a set of sample images are provided to each attendee. The format is intended to be flexible and you have the option of working individually or in groups on the assigned tutorials.

The data used in the tutorial consists of a complete set of images from different processing stages during an invasive plant species mapping project. These images were extracted from data sets used to map invasive plants at U.S. military bases by the CSTARS lab. The tutorial data set referenced in the lessons was created specifically to use with the tutorials provided in the manual.

The use of this tutorial assumes access to the ENVI software program 0 (Research Systems, Inc., Boulder, CO). The image classification techniques used to map the invasive plants for this project are performed using ENVI 4.. ENVI is used in the tutorial manual (Research Systems, 2002); it was modified and used with permission (RSI, 2003) to provide detailed instructions to allow an ENVI-novice to complete all the exercises in the manual. However, if you have ENVI, you also have the option, at any time of accessing the online help provided with the program.

During the course of two days, you will work through each of the image processing procedures that were used during the SERDP mapping project. The more basic procedures are covered with a full day of instruction on the first day. The second day will be a combination of practicing the techniques learned on the first day, field equipment demonstrations and presentations on some of the more advanced topics such as spectral unmixing and Mixture-Tuned Matched Filtering.

1.2 Invasive Plant Species Threat at U.S. Military Installations

The rapid spread of invasive plant species is causing irreparable damage to the natural resources on military installations. Current estimates indicate that in the U.S., invasive plants infest 70,000 hectares and cost more than $100 billion annually, causing severe economic losses and ecological degradation (Babbitt, 1998; Pimentel et al., 2000). Plant invasions are one of the most serious and rapidly accelerating ecological threats to land degradation and long-term environmental health faced by military environmental managers.

The significant role that ecosystem disturbance plays in facilitating the spread of invasive plant species creates a challenge for military resource managers, who must continue to fulfill the military mission, while reducing or mitigating impacts on ecosystem structure and function. Troop, vehicle, and other operational and training activities on military bases cause the types of landscape disturbances that facilitate the spread of invasive plant

species, while the inadvertent transmittal of invasive plants during military transport of people and equipment around the world exacerbates these problems. DOD, along with other federal agencies, has been mandated to develop proactive management techniques to protect native species and preserve the structure and function of natural ecosystems.

New remote sensing technologies, especially hyperspectral imagers, may improve monitoring and mapping capability to improve early detection and management of invasive plants. Our research addresses two of the objectives listed in the SON: 1) develop methods to inventory the spatial extent, and where applicable, the densities of established populations of invasive plants; and 2) develop methods to estimate the likelihood of invasive plant species expansions into and/or invasion of new habitats/areas on military bases.

1.3   Mapping with Remote Sensing

A hyperspectral methodology was developed for mapping invasive plants which has broad applicability to the military installations in this study and to management at other bases. To provide the broadest basis for assessing transportability, the invasive plant species selected have a range of growth characteristics (herbaceous, grasses, shrubs, trees) from military bases with different missions (Vandenberg AFB, Camp Pendleton, Yakima Training Center, Yuma Proving Ground, Ft. Benning, Aberdeen Proving Ground and Eglin AFB) in different climate zones (Northwest Great Basin semi-desert, southern California Coastal, Sonoran desert, Atlantic coast wetlands, Gulf coast wetlands, and southeastern hardwoods).

The tutorial addresses the unique spectral, spatial, and temporal characteristics of each invasive plant studied, to assess the applicability of using hyperspectral imaging (HSI) sensors for invasive plant species mapping. Commercial programs were used where appropriate, in order to facilitate technology transfer, but the tutorial also investigates more advanced techniques e.g. non-linear image analysis and other linked spatial-spectral analysis procedures like wavelet decomposition and multi-scale resolution when required for discrimination of the invasive plant species from the environmental background.

2   Data Acquisition and Fieldwork

2.1   Target Invasive Plant Species

The example invasive plants used in this tutorial were chosen by their respective military base personnel as important invasive plant species for their management. Nonetheless, not all invasive plants have potential to be mapped using remote sensing techniques. Because invasive plants occupy landscapes with native species it is not sufficient to merely detect the presence of vegetation as the indicator of invasive plant species

presence. Several criteria must be present to successfully identify the invasive plant species. First the invasive plant should occupy several pixels, preferentially clustered to improve the identification. The invasive plant species has to be there in sufficient amount to form a substantial part of individual pixels to be detected The exact fraction required depends on the relative spectral and spatial contrast between the invasive plant and the surrounding vegetation, with the greater the separation the less of the invasive plant needs to be present to detect its presence. Contrast can be improved if phenological differences exist between target invasive plants and/or native vegetation and if the invasive plant species have some aspect of unique biology, ranging from different growth forms (grasses vs. shrubs or trees), biogeochemistry (different proportions of pigments, water, and dry matter sufficient to characterize it) and/or plant and canopy density. Lastly, the pixel-to-pixel variability of the "background vegetation" should be low to avoid confusing the background vegetation with varying amounts of invasive plants in the pixels. This tutorial illustrates detection of invasive plants under a range of external conditions.

2.2   Hyperspectral Data Acquisition

Specific detail about the data acquired for each site is available in past annual SERDP reports and in this report in the results section below. This study acquired data from NASA's Advanced Visible Infrared Imaging Spectrometer (AVIRIS), measuring 224 spectral bands over the 400-2500nm interval and having a nominal wavelength resolution of 10nm.   Several flightlines of data were acquired at 3-4 m pixel resolution at each base.
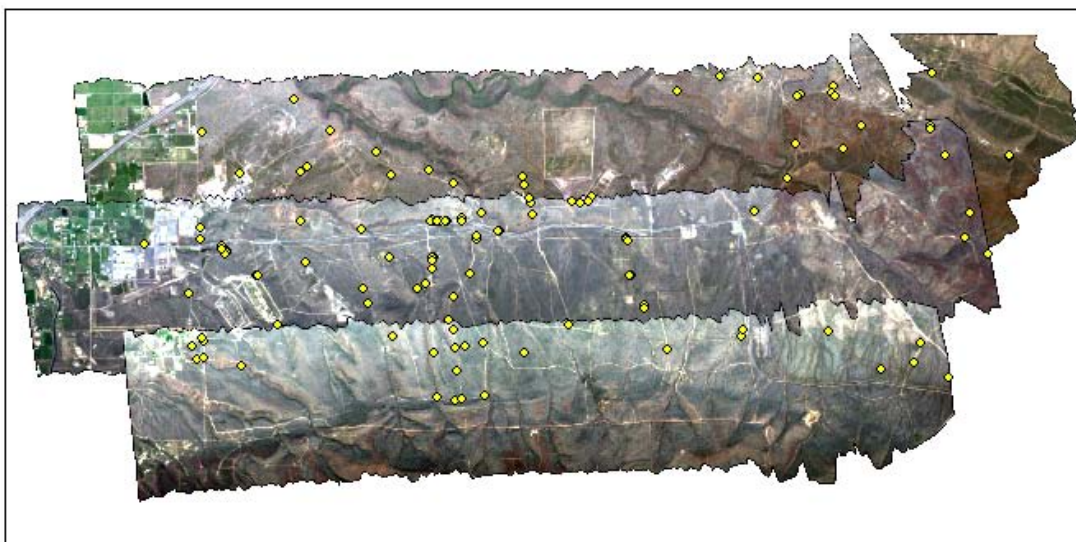

Figure 2-1: True color composite of three AVIRIS flightlines. GPS sampling locations are displayed as yellow dots.

2.3   Spectral Fieldwork

Field spectral measurements were acquired to meet two goals: (1) acquiring spectra of invasive plant species and native vegetation, soils, and other materials to use in image analysis, and (2) acquire spectra of "invariant" targets, i.e., those that are spectrally stable over time and can be used to improve the calibration of the data to surface reflectance. Examples of invariant targets are large (several pixels in size) cement or blacktop parking lots, bare dirt fields or roads, etc. The spectra are used to atmospherically calibrate the image data and in spectral unmixing. Locations of spectra and field notes are made along with GPS location.

Canopy spectra were measured using two field portable spectrometers, Analytical Spectral Devices (ASD) Spectrometer. The ASD measures in the visible, near-infrared (NIR) and short-wave infrared (SWIR) spectral regions from 0.4-2.5 nm.

Measurements were made in the nadir orientation with a 25º field of view at about 1 meter above the canopy or water surface in the case of submerged aquatic plants. A black shade cloth was draped over the edge of the boat to minimize solar reflection from the boat. Three sample points were taken at each location and each sample point was an average of ten spectra. Spectra were calibrated to surface reflectance by ratioing spectra to a horizontally leveled Spectralon panel (Labsphere, Inc., North Sutton, NH). Spectra were measured ± 2 hours from the time of the overflight in the morning and again in the afternoon (when the sun was at the same angle) to reduce solar angle effects and sun glint off the water.
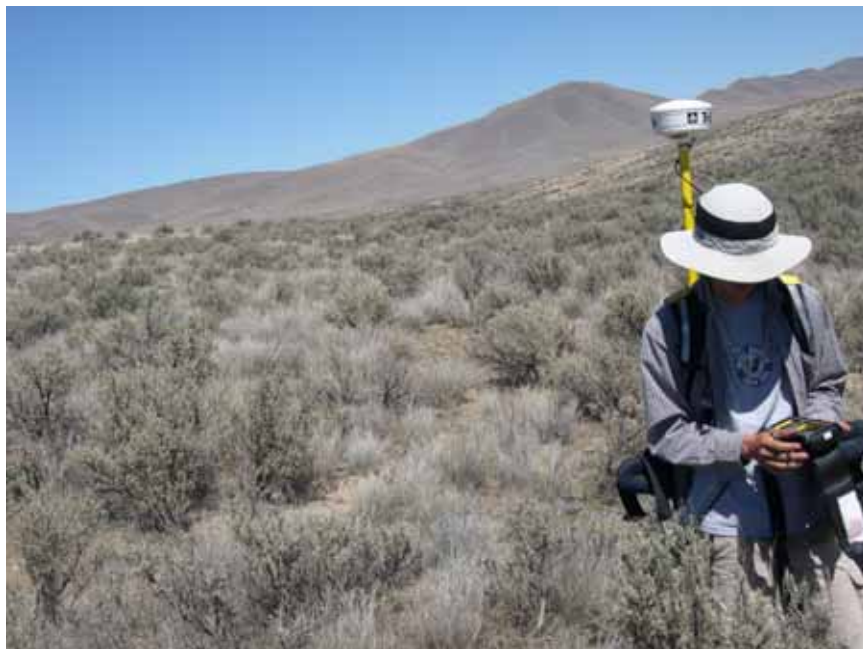
Figure 2-2: Spectral measurements taken during fieldwork in the shrubland community of Yakima Training Center.

2.3.1   Locate Ground Control Points

GPS locations of road intersections and other objects that can be unambiguously identified in the imagery were used as ground control points (GCPs). These known locations are used to more accurately georeference image data. If aerial photos are available, these can be used to identify potential ground control points prior to entering the field.

2.3.2   Collect Regions of Interest (ROIs)

Both during the initial field sampling and during the validation sampling, polygons of "homogeneous" materials are identified to be used to identify particular surface materials in the imagery. Typically ROIs are several pixels in extent and could be bare soil, road materials, invasive plant species, native vegetation (either community types or individual species). The goal is to identify as many materials as possible and limit the potential of having other materials in the image that could be confused with the invasive plants.

ROIs were collected primarily by using a Trimble Pro-XRS (Trimble, Inc., Sunnyvale, CA) to record the location of areas of interest. Range finder binoculars were used to record the locations of vegetation species located on the levees, on islands within the channels, and in areas that the airboats could not reach.

2.3.3   Create Spectral Library

Field spectra are examined and any bad data removed. Spectra become part of the library of materials for each base and are used to either (1) train a classifier to identify the material or (2) used to evaluate the "correctness" of a classification or (3) used to determine which "pure" materials (i.e., spectra of known objects) are detectable if they are the only materials in a pixel and (4) how likely the material is detectable when it is mixed with other materials at the site. For example, when a plant has partially senesced, it may spectrally look like a mixture of two other species.

a)   Calibration Targets

Calibration targets are the "invariant" targets that are sufficiently large that pure pixels of the material can be identified in the image. Ideally, both a dark target (e.g., a lake or large, deep water body or a dark material like asphalt) and a light target (e.g., a sand pit, beach, or cement structure) are needed to improve the calibration of the image to surface reflectance. Typically these are used after a radiative transfer model has been used to calibrate scattering and absorption by the atmosphere. These targets are used by

performing a regression between the "actual reflectance measured in the field and the reflectance retrieval of the pixels where the field measurements were made after the atmospheric radiative transfer model has been applied. The resulting set of linear equations (one for each waveband) is then applied to all pixels in the image. This step will eliminate a lot of artifacts in the image data, but if not done carefully (and with subsequent checking of other known ground targets), can introduce errors.

b)   Vegetation

Spectra of individual plant characteristics of each of the native plant communities and of the invasive plant species are measured in as many site locations as possible, given time constraints. Spectra of plants at different densities and stages in their life cycles should be collected to improve the spectral library. The goal is to identify as many conditions as realistically possible.

Sampling was undertaken to collect information on the invasive plants of concern in the SERDP as well as additional native and non-native species that are found in similar areas. The goal was to determine how well the various species found in similar niches throughout the military installations could be differentiated. The locations of the sample locations were random although consideration was given to areas that were easily accessible, had large monotypic areas of invasive plant species of interest that were not likely to move, patches that were not obscured by over hanging branches, and areas that were a mixture of different combinations of vegetation species.

At the same time field spectral signatures were collected on each boat, GPS readings of the locations were taken using a Trimble Pro-XRS with an accuracy of <1m. Site-specific characteristics were recorded such as percent cover, surrounding vegetation, and water clarity. Digital photographs of each sample area were taken.   A total of 224 spectral signatures from various species were collected, measured, and recorded with the ASD and GPS instruments

c)   Soils

Spectra of the dominant soil types that occur in each of the plant communities are acquired. Spectra are taken of large openings of bare soil where possible. Because the underlying soil reflectance beneath the plant canopy affects the pixel reflectance, this step increases the likelihood that the image data will be reproducible from the spectral library by mixing the plants, soils, and plant litter in various proportions.

2.3.4   Processing of field spectrum

Spectra from the field were input, managed, and analyzed in the Spectral Analysis and Management System (SAMS) (a proprietary package of CSTARS). SAMS was used to organize field spectra and delete noisy data. The spectral signatures of the three sample

points taken per species/location were averaged. SAMS was used to compare the spectra for each species and spectra between different plants of the same species. SAMS was used to determine appropriate endmember selection, compare pixel endmembers to field referenced spectral signatures, and begin a spectral library.

3   Overall Technical Approach

A combination of field knowledge of invasive invasive plant species and image analysis were used to develop maps of invasive plants for the bases. This approach has been to obtain preliminary information from the base GIS database on topography, road and construction layers, and any vegetation, soil or geology maps of the base. These were used in consultation with base personnel to develop an over flight plan for the base that included known locations of the target invasive plant species and areas of the same native vegetation that were not invaded by the invasive plants. Field sampling locations were chosen to be within the flightlines and include areas of invaded and non-invaded vegetation.

Following preliminary classification of the image data, a field validation survey was conducted to ascertain the accuracy of the proposed invasive plant species maps. Based on the new data, invasive plant species maps were evaluated for accuracy and then either updated and re-analyzed or accepted as meeting accuracy requirements.

The validated map is used in conjunction with the ecological analysis of the native-invasive plant species interactions to develop an environmental management model as shown in Figure 9 of the Final Report. A schematic of the hyperspectral image processing steps which are performed to achieve the validated invasive plant species map is provided in Figure 10 of the Final Report.
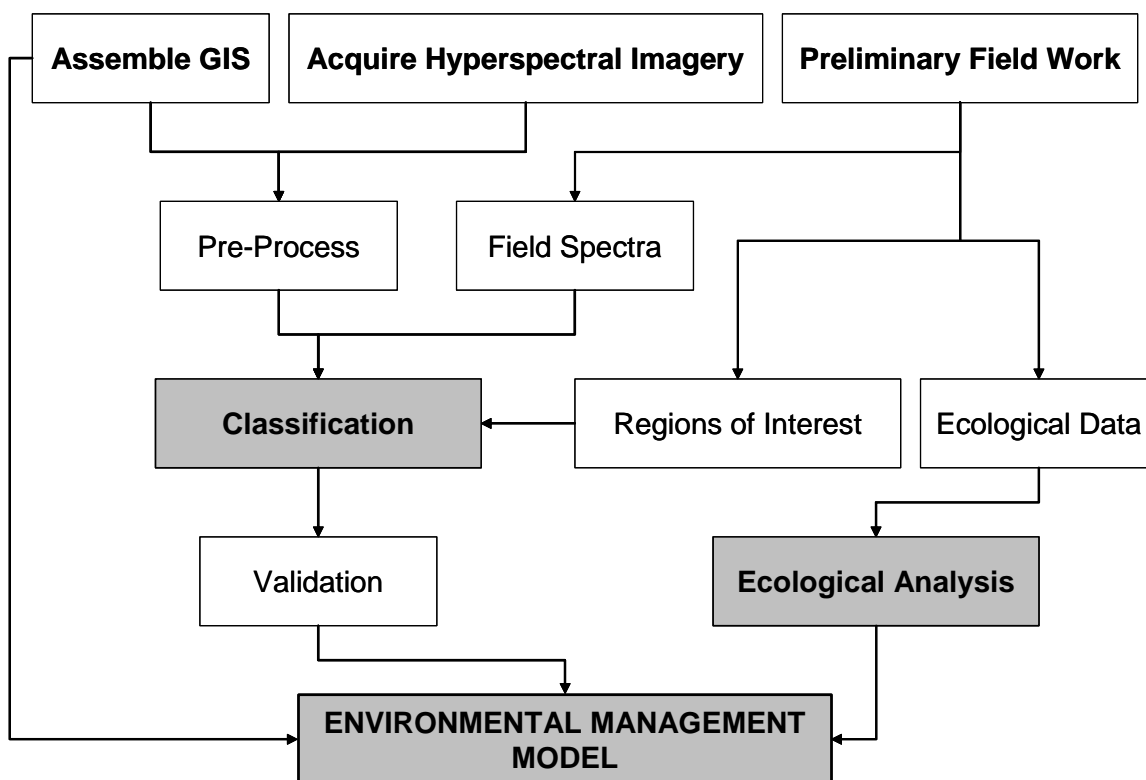
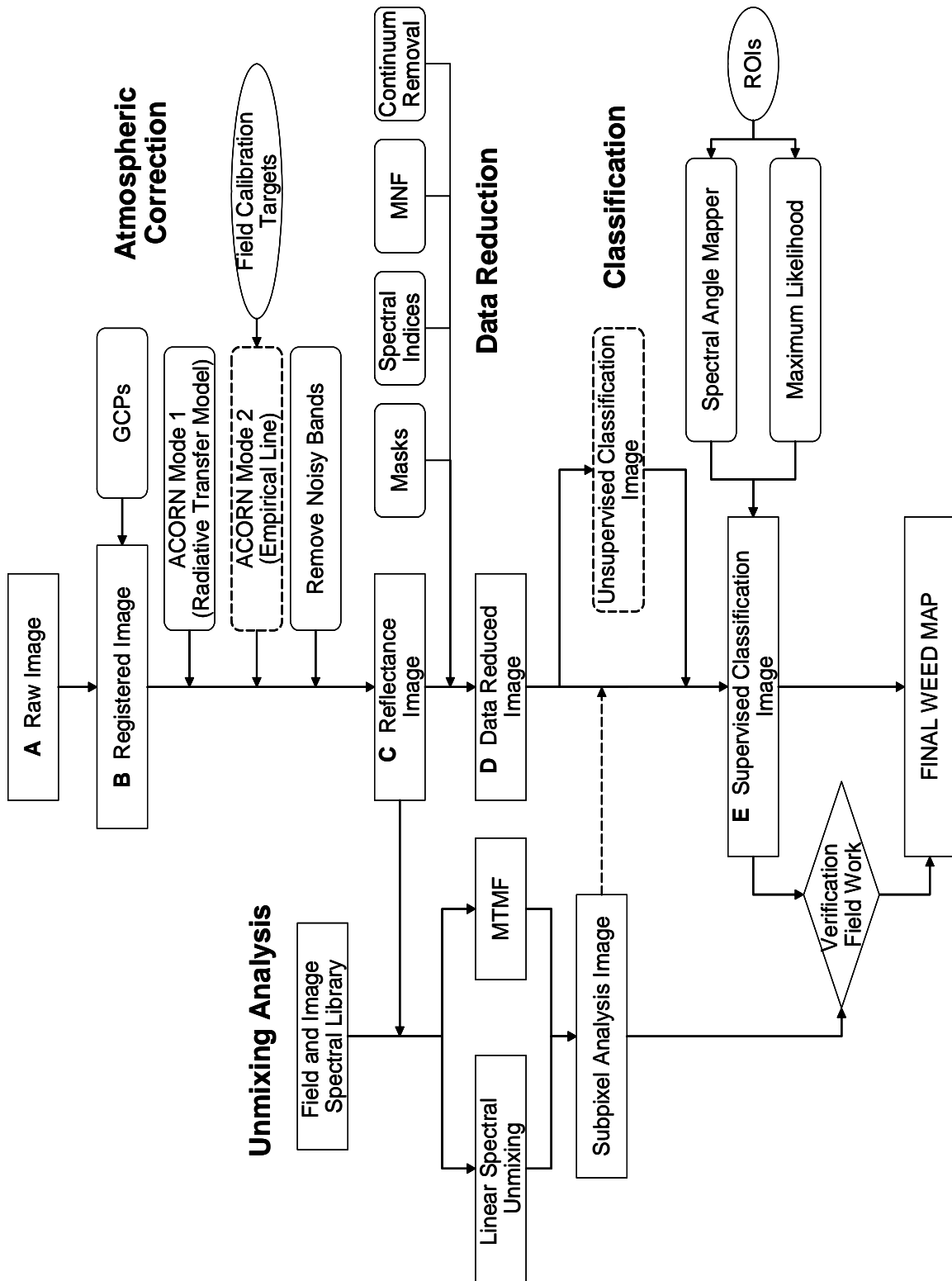**Figure 3-1: Schematic showing overall technical approach for the project.**

Figure 3-2: Schematic of hyperspectral image processing steps.

4   Imagery Pre-Processing

There are several steps that are required to be performed on hyperspectral imagery in order to fully utilize the information content of the data.

4.1   Image Registration

During image registration, the AVIRIS imagery is registered to the base layer of the GIS. The base layer consists of pre-existing aerial photos or digital maps of the study area. Often the field ground control points (GCPs) are used to fine-tune the registration of the base layer first.

JPL also geocorrected the ten images using proprietary software that uses sensor position and orientation data collected simultaneously with the image to calculate the position of each pixel in the image. There was still some geometric distortion within each flight line. The flight lines were further georeferenced to USGS 1 meter digital orthophotoquads.

Lesson 1: Building Spectral Library

CD-ROM: *SERDP Training & Data CD*

Required Files: `endmembers_reflectance.sli`
                         `Ykm_spectral_library.sli`
                         `usgs_veg.sli`

The **Spectral Libraries** module in ENVI is used to build and maintain personalized libraries of material spectra, and to access several public domain spectral libraries. ENVI spectral libraries are stored in ENVI image format, with each line of the image corresponding to an individual spectrum and each sample of the image corresponding to an individual spectral measurement at a specific wavelength. The ENVI spectral libraries can be displayed and/or enhanced using the standard ENVI image display and analysis routines.

Opening Spectral Libraries

Use **Spectral Library Viewer** to open and view the contents of a spectral library and to plot individual library spectra.

Select **Spectral** → **Spectral Libraries** → **Spectral Library Viewer**.

In the **Spectral Library Input File** dialog, click **Open Spec Lib**.

Select `endmembers_reflectance.sli`

Click **OK**.

Plotting Library Spectra

1. Open **Spectral Libraries** and use the **Spectral Library Viewer** dialog to plot the spectra.

2. In the **Spectral Library Viewer** plot window, click on an individual spectrum in the **Spectral Library Viewer** dialog.

3. Click on multiple spectrum names to plot more than one spectrum.

4. Repeat steps 1 & 2 with `Ykm_spectral_library.sli` and `usgs_veg.sli`.

Building Spectral Libraries (Optional)

Use **Spectral Library Builder** to create ENVI spectral libraries from a variety of sources including ASCII files, spectral files produced by ASD spectrometers, other spectral libraries, ROI means, and spectral profiles and plots. The collected spectra are automatically resampled to an input wavelength space using full-width-half-maximum (FWHM) information, if available.

Inputting Wavelength Data

1. Select **Spectral** → **Spectral Libraries** → **Spectral Library Builder**.

2. In the Spectral Library Builder dialog, select the wavelength set for the new library from either an **Input Data File** (ENVI image file) or from an Input ASCII File.

3. In the **Output Wavelength** dialog, select the input file containing the wavelengths and optional FWHM values using standard selection procedures.

4. If you select **Input Data File**, the wavelengths and FWHM values (if present) are read from the ENVI header file.

5. If you select **Input ASCII File**, the column that contains the wavelength values and FWHM values (if present) must be selected.

6. Click **OK**. The **Spectral Library Builder** dialog appears.

Collecting Spectra

Use the **Spectral Library Builder** dialog to collect endmember spectra from a variety of sources. All spectra are automatically resampled to the selected wavelengths.

The methods for collecting spectra in the **Spectral Library Builder** dialog are the same as those for collecting spectra in **the Endmember Collection** dialog.

Writing Spectral Libraries

Select an output option to build the spectral library with the collected spectra.

1. In the **Spectral Library Builder** dialog, select **File** → **Save Spectra As** → **ASCII** to create an ASCII file containing all of the spectra.

2. Or select **File** → **Save Spectra As** → **Spectral Library** to create an ENVI spectral library either as an output file or in memory.

Lesson 2: Image Registration

CD-ROM: *SERDP Training & Data CD*

Required Files: `Ykm3_TM.img`
`Ykm3_DOQQ.img`

Generated Files: `Ykm3_TM_warp.img`
`Ykm3_gcp.pts`

Designed as a starting point for users doing image registration, this section provides basic information about georeferenced images in ENVI and Image-to-Image Registration

Georeferenced Images in ENVI

ENVI's image registration and geometric correction utilities allow you to reference pixel-based images to geographic coordinates and/or correct them to match base image geometry. GCPs are selected using the full resolution (Main Image) and Zoom windows for both image-to-image and image-to-map registration. Coordinates are displayed for both base and uncorrected image GCPs, along with error terms for specific warping algorithms.

Warping can be performed using Resampling, Scaling and Translation (RST), polynomial functions (of order 1 through *n*), or Delaunay triangulation. Resampling methods supported include nearest-neighbor, bilinear interpolation, and cubic convolution. The "nearest neighbor" uses the original data values to fill in the additional pixels created during the registration warping and is usually the preferred method to maintain the data quality of the image.

Comparison of the base and warped images using ENVI's multiple Dynamic Overlay capabilities allows quick assessment of registration accuracy. As a general guideline, the total RMS error from the registration warping process is maintained around 0.5 pixels.

Image to Image Registration

For this tutorial, a georeferenced aerial photo and a text file of GCPs were used to register an AVIRIS image. Ideally, GCP's are taken from the field, converted into a point file or vector layer, then used as a base layer to which an image can be warped. However, georeferenced images also can be used as a substitute for field data.

Open and Display Georeferenced Aerial Photos

1. In ENVI's main menu select **File** → **Open Image File**.

2. In the **Enter Data Filename** dialog and select `Ykm3_DOQQ.img`. Note: ENVI does not always display the file extension in this dialog.

3. Click **Open**.

4. In the **Available Bands List** click **Gray Scale** radio button, then click **Load Band** button to display the orthophotoquad image.

<u>Open and Display the Simulated Landsat TM Image</u>

Because registering AVIRIS imagery with 224 bands requires considerable computing time and disk space, the AVIRIS data was filtered to provide a broadband image simulating the spectral output of the Landsat Thematic Mapper (TM). This smaller spectral subset will be used for this exercise.

To open the simulated Landsat TM image:

1. In ENVI's main menu, select **File → Open Image File**.

2. In the **Enter Data Filename** dialog navigate to the data directory and select `Ykm3_TM.img`.

3. Click **OK**.

4. In the **Available Bands List** dialog, click radio button **RGB Color** and select the first three bands- 1, 2 & 3.

5. Click **New Display**.

6. Click **Load RGB**.

<u>Ground Control Points (GCPs)</u>

Select GCPs by comparing the base image (aerial photo) with the warp image (simulated Landsat TM image) and identifying landscape features visible in both images. Road intersections, buildings, and waterways make good choices. Mark these features and use them as GCP pairs to register the warp image to the base image.

1. In ENVI's main menu select **Map → Registration → Select GCPs: Image to Image**.

2. In the **Image to Image Registration** dialog choose `Ykm3_DOQQ.img` (Display #1) as the base image and `Ykm3_TM.img` (Display #2) as the warp image.

3. Click **OK.**

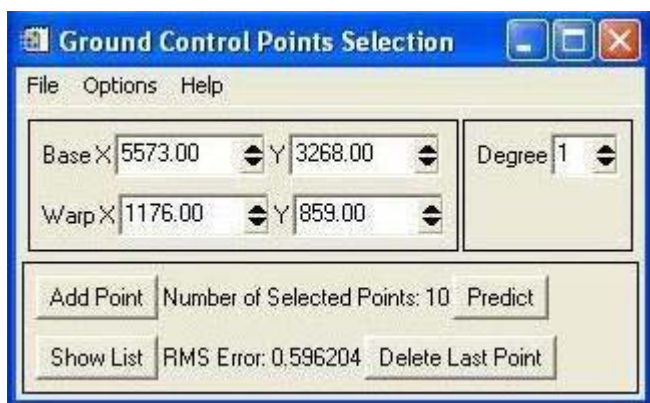The Ground Control Points Selection dialog appears.

Figure 4-1: Ground Control Points Selection dialog for Image to Image Registration

4. Explore the image and zoom in to several different features to help you identify those that make good GCPs. When you have found a good point, click **Add Point**. The GCP is now displayed on both the base (`Ykm3_DOQQ.img`) and warp (`Ykm3_TM.img`) images.

5. Select at least 4 GCP points.

6. Click **Show List**. The **Image to Image GCP List** dialog appears. Examine the base map coordinates, the actual and predicted image coordinates, and the RMS error.

7. After the desirable level of RMS is reached, click **File** → Save **GCPs to ASCII**…; enter a name `Ykm3_gcp.pts`.
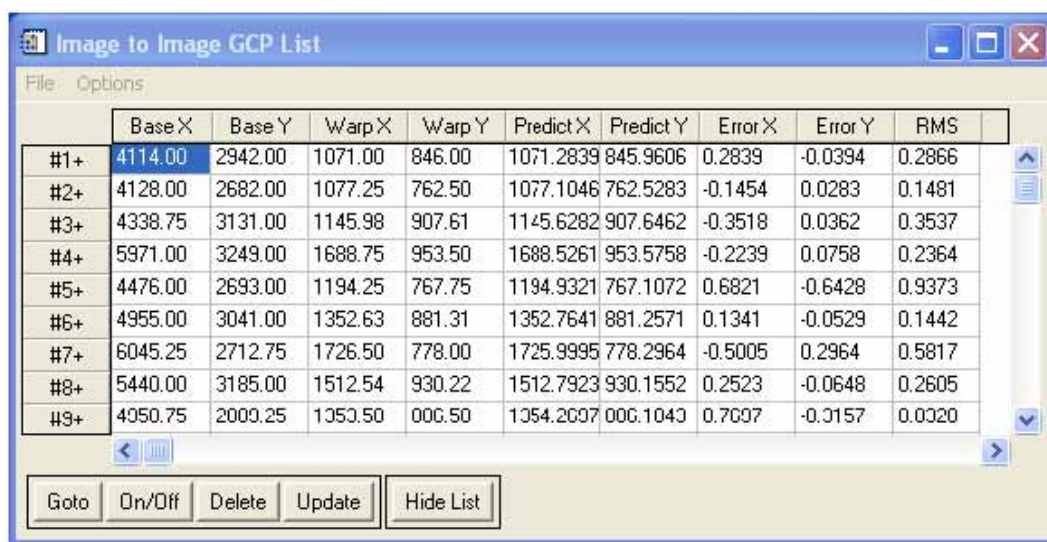
Figure 4-2: Image to Image GCP List dialog for image to map registration

Warping

1. Now that GCPs have been established, georectify the Landsat TM image.

2. **GCP Selection dialog → Options → Warp File**.

3. Select `Ykm3_TM.img` as the Input File and click OK.

4. Save as `Ykm3_TM_warp.img` and click OK.

5. In the **Available Bands List** dialog, select `Ykm3_TM_warp.img`, click **Display → New Display → Load Band**.

Note that the new image has been rotated so that its coordinates match `Ykm3_DOQQ.img`.

6. In Display #1, select **Tools → Link → Link Displays**.

7. Select **No** for the old, unwarped, Landsat TM image `Ykm3_TM.img` in Display #2), click **OK**.

8. Close the unwarped Landsat TM image.

The new image (`Ykm3_TM_warp.img`) and the reference image (`Ykm3_DOQQ.img`) are now linked. Dragging the zoom box in one image window will move it in the linked window.

`Ykm3_TM_warp.img` is now registered. Optional exercises on editing map information and Image to Map registration are provided on the following pages.

Edit Map Info in ENVI Header (Optional)

1. In the **Available Bands List**, right click on the Map Info icon under `Ykm3_DOQQ.img` and select **Edit Map Information** from the shortcut menu.

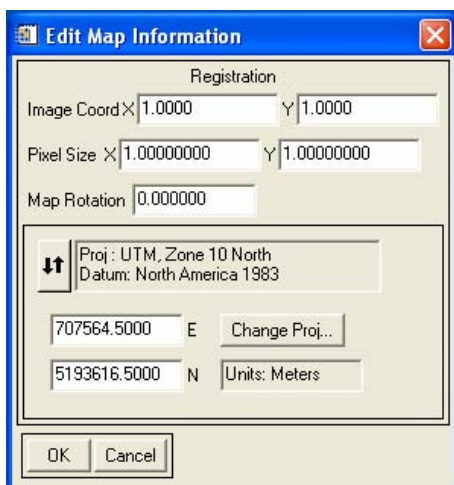The **Edit Map Information** dialog appears.

Figure 4-3: The Edit Map Information dialog

This dialog lists the basic map information used by ENVI in georeferencing. The image coordinates correspond to the Magic Pixel used by ENVI as the starting point for the map coordinate system. Because ENVI knows the map projection, pixel size, and map projection parameters from the header information and the map projection text file, it can calculate the geographic coordinates of any pixel in the image. Enter coordinates as either map or geographic (latitude/longitude).

2. Click next to the **Projection/Datum** field to display coordinates.

3. Click **DMS** or **DDEG** to toggle between Degrees-Minutes- Seconds, and Decimal Degrees, respectively.

4. Click **Cancel** to exit the **Edit Map Information** dialog.

Cursor Location/Value

1. Open a dialog box that displays the location of the cursor in the **Main Image**, **Scroll**, or **Zoom** windows.

2. From the **Main Image** window menu bar, select **Tools → Cursor Location/Value**. Or open this dialog from the ENVI main menu by selecting **Window → Cursor Location/Value**.
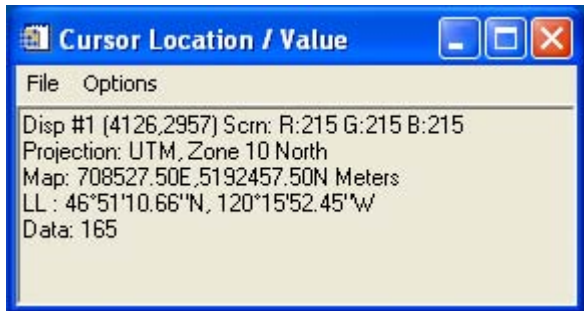

Figure 4-4: The Cursor Location dialog displays the pixel and georeferenced coordinates for georeferenced images

Note that both pixel and projected coordinates are given.

3. Move the cursor around the image and examine the coordinates for specific locations. Note the relation between map coordinates and latitude/longitude.

4. Select **File → Cancel** to end the dialog.

4.2 Atmospheric Correction

The primary objective of atmospheric correction is to simulate the effect of the atmosphere on the image pixels and to remove it. During this process, the water absorption bands at 820, 940 and 1140 nm are removed from the spectral data set and the input radiance spectra are converted to apparent surface reflectance.

Sensor radiance data were converted to apparent reflectance by the JPL Corporation using HYCORR software. This is a program for converting radiance AVIRIS images to apparent surface reflectance. Two levels of processing occur. The first is compatible with ATREM3 processing. The more advanced level has an ATREM pass followed by an EFFORT polishing pass (spectral enhancement) to remove systematic ATREM errors. EFFORT processing was used on this data set. The derived surface reflectance image has been scaled by 10,000 (reflectance *10000) and the wavelengths are displayed as nm.

There are several other commercial software packages which will atmospherically correct hyperspectral data. In this tutorial, a program called ACORN (Analytical Imaging and Geophysics) is used. The correction algorithms in Mode 1 of ACORN, accept basic physical parameters regarding overflight conditions (time of day, day of year, altitude of plane) as inputs to a radiative transfer model based on the MODTRAN program. The radiative transfer model extrapolates a hypothetical "standard" atmosphere based on this information and subtracts the hypothetical moist atmosphere from the image data on a pixel-by-pixel basis to create a reflectance image. A byproduct from this analysis is that three additional images are created, an image of the density of water vapor in equivalent thickness of liquid water and an image of the distribution of liquid water in the vegetation canopy in the image. A third image is possible, that is, if ice or snow is present in the image, it can be mapped. The basis for this discrimination is that the maximum absorption band for water is shifted about 40 nm (4 bands) for water in vapor, liquid and solid phases.

If field calibration targets are available, Modes 2 and 3 use these targets to fine-tune the atmospheric calibration to surface reflectance that matches actual field conditions.

Lesson 3: Atmospheric Correction

CD-ROM: *SERDP Training & Data CD*

Required Files: `Ykm3_raw.img`
                               `Ykm3_reflectance.img`

Correcting spectral images for atmospheric conditions using Atmospheric COrrection Now (ACORN) software from Analytical Imaging and Geophysics LLC is covered in this section.

Software

ACORN uses radiative transfer calculations and the measured, calibrated hyperspectral data, to deduce a subset of the atmospheric properties present in the hyperspectral data. These derived atmospheric properties are used in conjunction with modeled atmospheric properties to estimate and remove the effect of atmospheric conditions on the hyperspectral data. With an input of calibrated hyperspectral radiance data, ACORN produces an output of apparent surface reflectance. The hyperspectral data must be spectrally and radiometrically calibrated to use ACORN. The ACORN user controls the strategy for water vapor estimation, artifact suppression, and visibility constraint and estimation.

Several types of files are needed to effectively use the ACORN software. Because these files can take a substantial amount of time to generate, so the output images are provided for comparison.

Field Data

Field data from spectrally flat calibration targets visible in the images provide additional calibration information. Spectral data need to be taken from these targets 2 hours before or after solar noon and as close as possible to (preferably on) the same day the images are taken. Also, recorded observations about atmospheric conditions at the time field data are collected can be entered into the calibration software. ACORN converts the image data from radiance to reflectance.

At least one pair of targets, one light and one dark, is required to execute a Mode 2 calibration in ACORN. Multiple targets are preferred and, more importantly, they should be spread throughout the image. ACORN uses the differences between the field spectral data and the raw images to develop algorithms that correct all of the pixels in the raw images. If both of these targets are visible in multiple flightlines, use them to cross-calibrate other flightlines.

View Atmospheric Correction Results

1. In ENVI's main menu select **File → Open Image File.**

2. In the **Enter Data Filename** dialog, select `Ykm3_reflectance.img` and load into **Display #1**.

3. Also open `Ykm3_raw.img`, display the same bands in **Display #2**.

4. Select **Tools → Link → Link Displays** from any image display.

5. Select **Yes** for both displays.

6. Right-click on the image in the first display window, select Z-profile. This resulting spectral profile is for the pixel in the cross hairs of the zoom window.

7. Repeat for the other image and compare the differences.

Note how in the ACORN processed `Ykm3_reflectance.img` spectral profile, the main water absorption bands are removed, and the spectral profile is much smoother than the other image.

5   Image Classification

To classify an image is to assign each pixel in the image to a class. The spectral band values for each pixel in an image forms a cloud of points when plotted in multi-dimensional space. A classifier is a computer algorithm that takes the data cloud of points and groups the data into clusters or classes.

The assumption used by classifiers is that pixels of like materials will plot close together in the data cloud, and that the closest cluster to an individual data point will consist of other pixels of the same category. This is demonstrated schematically in Figure 5-1, which shows a data cloud in 2-D image space.
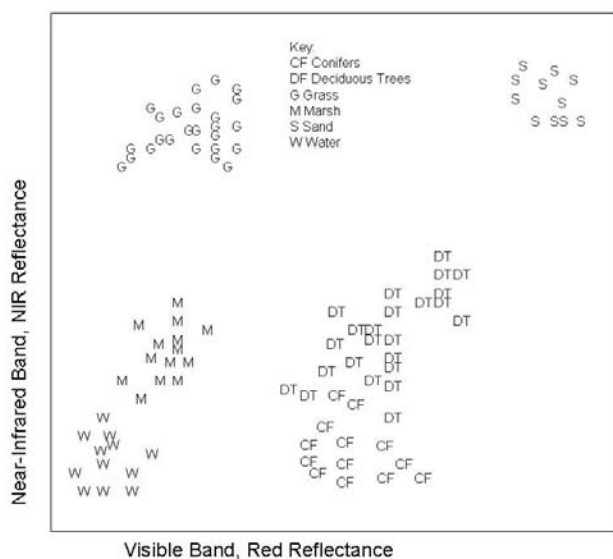

Figure 5-1: Image pixels grouped into classes in a 2D scatter plot

There are two major types of classification: supervised and unsupervised. Supervised classification uses thematic classes defined by the user (invasive plant species, native vegetation, roads, etc.) and unsupervised classification uses spectral classes that are statistically defined by the computer classifier to group pixels based on their tendency to cluster in the data cloud. Both types of classifiers are used in this tutorial.

5.1   Unsupervised

An unsupervised classification is often used as a preliminary stage prior to supervised classification. The unsupervised classifier delineates the clusters within the image based purely on their spectral properties, which may or may not represent classes of interest to the user. However, it does provide an independent estimate of the number and types of independent classes in the data.

The spectral classes found by the computer are then compared to the thematic classes defined by the user to fine-tune the definition of the thematic classes. If there is more than one spectral grouping within a thematic class (i.e., stressed vs. unstressed vegetation, dense vs. sparse canopy forest, etc.), then the thematic classes may be further broken down into finer categories. On the other hand, if several thematic classes fall within the same spectral grouping, it may lead the user to combine categories to maintain a more cohesive spectral class.

Unsupervised classification can be very computer intensive since the computer is considering all possible data groupings. Therefore, it often requires data reduction techniques (such as covered in Section 6) prior to running on large data sets such as AVIRIS.

Lesson 4: Unsupervised Classification

CD-ROM: *SERDP Training & Data CD*

Required Files: `Ykm3_reflectance.img`
                                      `Ykm3_NDVI_mask.img`
                                      `Ykm3_vegtype.img`
                                      `Ykm3_DOQQ.img`
                                      `Ykm3_gcp.pts`

Generated Files: `Ykm3_isodata.img`
                                      `Ykm3_isodata_warp.img`
                                      `Ykm3_Kmeans.img`
                                      `Ykm3_Kmeans_warp.img`

Use **Unsupervised** Classification to cluster pixels in a data set based on statistics only, without any user-defined training classes. The unsupervised classification techniques available are Isodata and K-Means.

Using Isodata Classification

Isodata unsupervised classification calculates class means evenly distributed in the data space and then iteratively clusters the remaining pixels using minimum distance techniques. Each iteration recalculates means and reclassifies pixels with respect to the new means. Iterative class splitting, merging, and deleting are done based on input threshold parameters. All pixels are classified to the nearest class unless a standard deviation or distance threshold is specified, in which case some pixels may be unclassified if they do not meet the selected criteria. This process continues until the number of pixels in each class changes by less than the selected pixel change threshold or the maximum number of iterations is reached.

1. **File → Open Image File →** `Ykm3_reflectance.img` select mask band `Ykm3_NDVI_mask.img`.

2. Select **Classification → Unsupervised → Isodata**

3. In the **Classification Input File** dialog, select `Ykm3_reflectance.img`.

The Isodata Parameters dialog appears. Options include input of a range for the number of classes to be defined, the pixel change threshold (0 - 100%), the maximum number of iterations to be used to separate the classes, splitting, merging, and deleting classes thresholds, and optional distance thresholds.
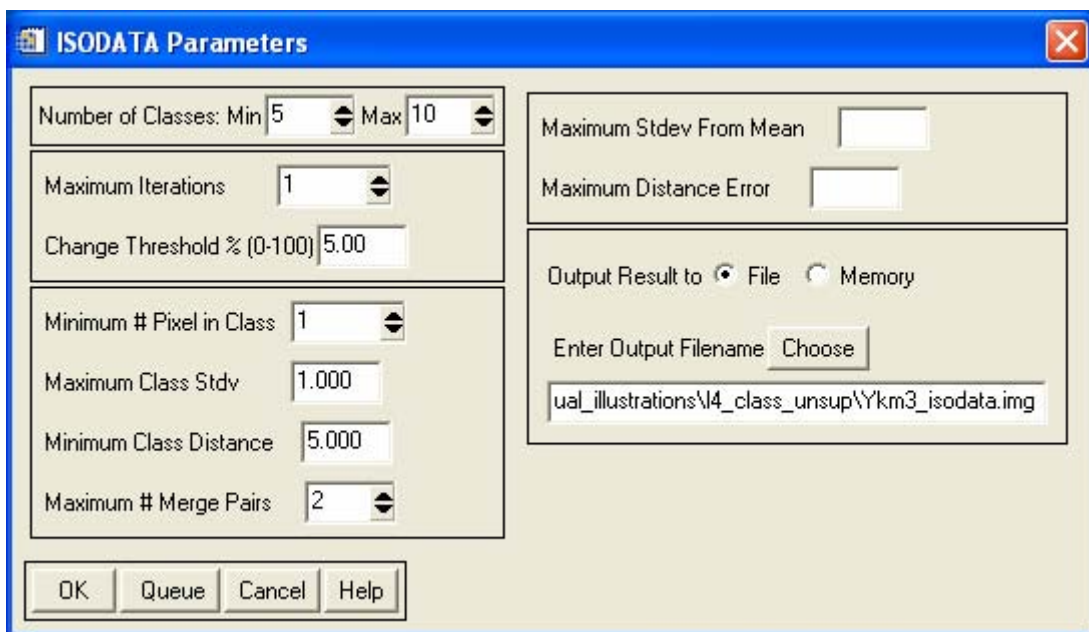
Figure 5-2: ISODATA parameters dialog box

4. Use the default values for steps 5 through 10.

5. Enter the minimum and maximum number of classes to be defined. A range for the number of classes is used because the Isodata algorithm splits and merges classes based on input thresholds and does not keep a fixed number of classes.

6. Enter the maximum number of iterations and a change threshold (0-100%) in the appropriate text boxes.

The change threshold is used to end the iterative process when the number of pixels in each class changes by less than the threshold. The classification will end when either this threshold is met or the maximum number of iterations has been reached.

7. Enter the minimum number of pixels needed to form a class in the appropriate text box.

If there are fewer than the minimum number of pixels in a class, then that class will be deleted and the pixels placed in the class(es) nearest to them.

8. Enter the maximum class standard deviation (in DN) in the **Maximum Class Stdv** text box. If the standard deviation of a class is larger than this threshold, the class is split into two.

9. Enter the minimum distance (in DN) between class means and the maximum number of merge pairs in the appropriate text boxes.

If the distance between class means is less than the minimum value entered, the classes will be merged. The maximum number of class pairs to be merged is set by the maximum number of merge pairs parameter.
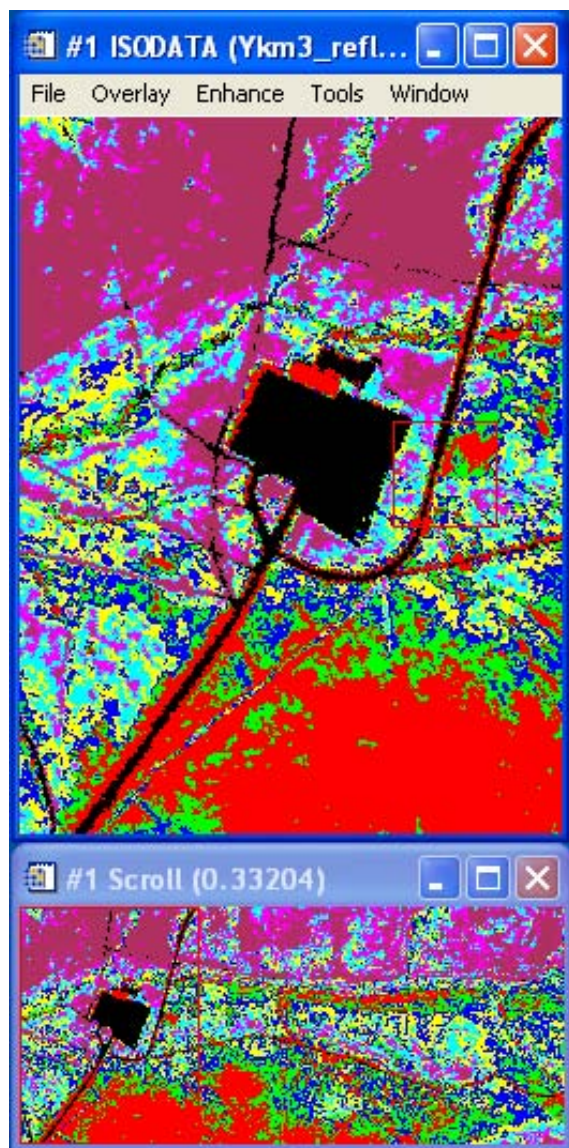
10. Set the optional standard deviation to use around the class mean and/or the maximum allowable distance error (in DN) by entering the values in the **Maximum Stdev From Mean**: or **Maximum Distance Error**: text boxes, respectively.

If both of these optional parameters are entered, the classification uses the smaller of the two to determine which pixels will be classified. If neither parameter is entered, then all pixels will be classified.

11. Select output to **File** (Ykm3_isodata.img).

12. Click **OK** to start the Isodata classification.

Statistics are calculated for each band of the image, and a status window displays the progress of the operation. The status bar cycles from 0 to 100% for each iteration of the classifier. The statistics for the initial class seeds are computed with a skip factor of 2.5 for both the sample and line directions.

Figure 5-3: Example of Isodata results for a subset of Yakima region.

Using K-Means Classification

K-Means unsupervised classification calculates initial class means evenly distributed in the data space and then iteratively clusters the pixels into the nearest class using a minimum distance technique. Each iteration recalculates class means and reclassifies pixels with respect to the new means. All pixels are classified to the nearest class unless a standard deviation or distance threshold is specified, in which case some pixels may be unclassified if they do not meet the selected criteria. This process continues until the

number of pixels in each class changes by less than the selected pixel change threshold or the maximum number of iterations is reached.

1. Select **Classification** → **Unsupervised** → **K-Means.**

2. Select `Ykm3_reflectance.img` file.

3. The **K-Means Parameters** dialog appears. Options available include selection of the number of classes to be defined by the clustering procedure, the pixel change threshold (0 - 100%), the maximum number of iterations to be used to separate the classes and optional distance thresholds. The default values can be used for steps 4 through 6.

4. Enter the desired number of classes and maximum number of iterations in the appropriate text boxes.
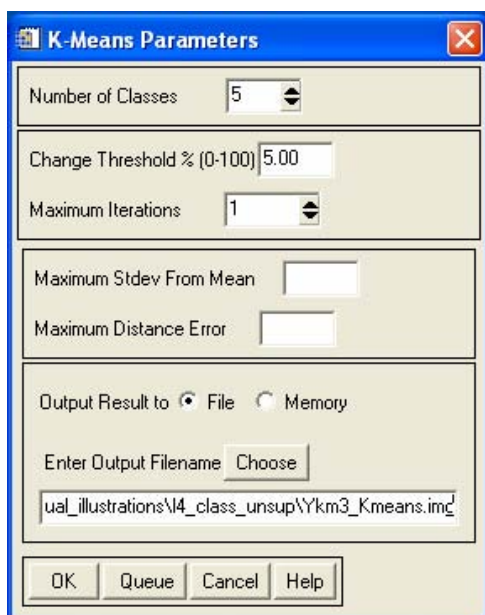
5. Enter a change threshold (0-100%) which will be used to end the iterative process when the number of pixels in each class changes by less than the threshold. The classification will end when either this threshold is met or the maximum number of iterations has been reached.

6. Set the optional standard deviation to use around the class mean and/or the maximum allowable distance error (in DN), enter the values in the **Maximum Stdev From Mean**: or **Maximum Distance Error**: text boxes, respectively.

Figure 5.4: K-means Parameters dialog

If both of these optional parameters are entered, the classification uses the smaller of the two to determine which pixels will be classified. If neither parameter is entered, then all pixels will be classified.

7. Select output to **File** (`Ykm3_Kmeans.img`).

8. Click **OK** to start the K-Means classification.

Statistics are calculated for each band of the image, and a status displays the progress of the operation. The status bar cycles from 0 to 100% for each iteration of the classifier.

The statistics for the initial class seeds are computed with a skip factor of 2.5 for both the sample and line directions.
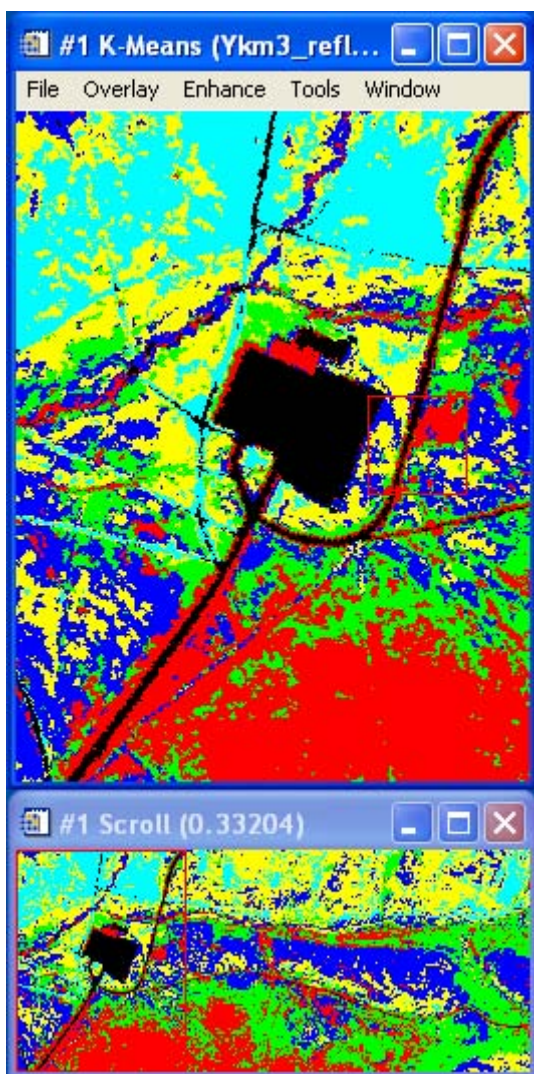

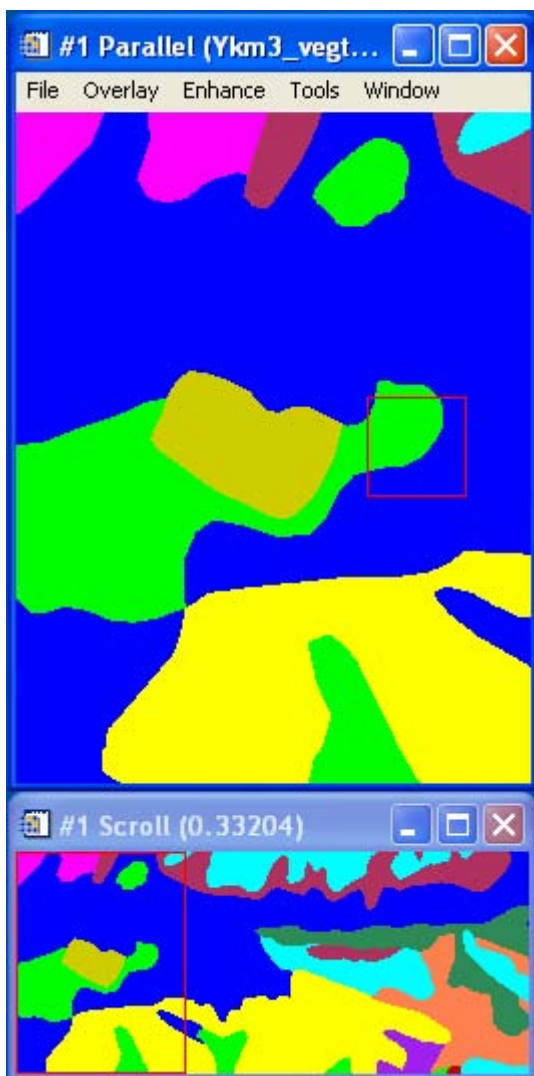*Figure 5-5: K-Means results for a subset of Yakima region.*

Figure 5-6: Vegetation type (Habitat suitability) map for the same subset of Yakima region.

Both Isodata and K-means classification results can be compared with the reference map of vegetation type `Ykm3_vegtype.img` or DOQQ layer to see if the classes selected by the computer match what the user can determine visually. Since the classification was done on the unregistered reflectance image (in order to preserve spectral characteristics of the data), and the reference image `Ykm3_vegtype.img` is georegistered, this data has to be geometrically correct using the `Ykm3_Isodata.img` and `Ykm3_Kmeans.img` It should be done as described in **Lesson2** of this **Manual**. DOQQ layer should be used as a base image, and isodata and K-means classification results should be used as warp images. The set of ground control points collected and saved in **Lesson 2** should be downloaded selecting **File → Restore GCPs from ASCII…** in the **Ground Control Points Selection** dialog.
Name output files `Ykm3_isodata_warp.img` and `Ykm3_Kmeans_warp.img`.

Now all three files: `Ykm3_isodata_warp.img`, `Ykm3_Kmeans_warp.img` and `Ykm3_vegtype.img` can be opened in three displays and linked to visually compare the unsupervised classification patterns with the reference map.

5.2   Supervised

Supervised classification uses spectral clusters defined by the user to assign pixels. The user determines the spectral properties of the class by defining an ROI within an image and extracting the spectral information for that group of pixels. The types of classes defined by the user depend on the themes of interest within the image. The major thematic classes often used for invasive plant species mapping are:

- invasive plant species vs. background
- primary vegetation groupings
- disturbance

The first level of classification is between invasive plant species and native vegetation (binary classification). This provides us with a map of the status of invasion at the time of overflight.

The next level of classification is to map the primary components of the major ecosystems present at each base. This type of classification groups the image into the major vegetation systems and distinguishes betweens levels of invasion. This component allows us to look at invaded vs. ecologically similar non-invaded areas and supply information for risk assessment modeling. The modeling can then help delineate the site conditions and locations in the study area vulnerable to invasion.

Another type of classification useful for this project is to determine levels of disturbance within the image. This is done by using GIS layers to help find disturbance endmembers within the image such as roads and trails, buildings, patches of erosion, etc. Since disturbance is usually a major factor in the spread of most invasive plants, this type of classification is an important information layer in the risk assessment modeling.

There are two main types of supervised classifiers used in this project, Spectral Angle Mapper and Maximum Likelihood (covered in the next couple of sections) and a couple of more advanced techniques covered in Section 7.

5.2.1   Spectral Angle Mapper

This description has been adapted from the ENVI tutorial manual (Research Systems, 2002).

The Spectral Angle Mapper (SAM) is an automated method for comparing the spectra of image pixels to the spectra of reference endmembers (CSES, 1992; Kruse *et al.*, 1993). SAM treats the spectra as vectors and determines the similarity between the unknown and the reference pixels by calculating the "spectral angle" between them.

A simple example is shown in Figure 5-7 which uses a reference spectrum (material A) and an unknown spectrum (material B) from a data set of *n* bands. The reflectances of each material can be plotted in a 2-D scatter plot for each combination of bands in the data set. A vector is then extended from the origin to the point of I-J reflectance for each material. The 2-band spectral angle for the reference and unknown pixels in the I-J data space is the angle between the two vectors. The overall spectral angle, α, can be thought of as equivalent to the sum of the angles for each 2-band combination of the *n* bands.
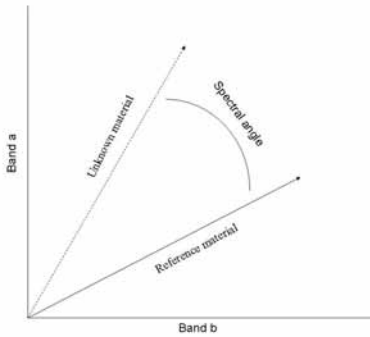

Figure 5-7: Two-dimensional example of the Spectral Angle Mapper

The color of a material is defined by the direction of its unit vector. The length of the vector relates only to how fully the pixel is illuminated. Because SAM uses only the direction of the spectra, and not the length, all possible levels of illumination are treated equally. Poorly illuminated pixels fall closer to the origin, but still lie along the same vector as brightly illuminated pixels. Notice that the angle between the vectors is the same regardless of the length.

The SAM algorithm generalizes this geometric interpretation to *n*-dimensional space. SAM determines the similarity of an unknown spectrum *t* to a reference spectrum *r*, by applying the following equation (CSES, 1992):

$$\alpha = \cos^{-1}\left(\frac{\rho_t \cdot \rho_r}{\|\rho_t\| \cdot \|\rho_r\|} \right)$$

which also can be written as:

$$\alpha = \cos^{-1}\left(\frac{\sum\limits_{i=1}^{n} t_i r_i}{\left(\sum\limits_{i=1}^{n} t_i^2\right)^{\frac{1}{2}} \left(\sum\limits_{i=1}^{n} r_i^2\right)^{\frac{1}{2}}} \right)$$

where *n* equals the number of bands in the image.

For each reference spectrum (endmember) in the analysis of a hyperspectral image, the spectral angle α, is determined for each image spectrum (pixel). This value, in radians, is assigned to the corresponding pixel in the output SAM image, with one output image for each endmember. The derived spectral angle maps form a new data cube with the number of bands equal to the number of reference spectra used in the mapping. The user specifies a threshold angle which is used to determine the maximum allowable angle difference between the reference and the unknown. If α exceeds the threshold angle, the unknown will not be placed in the reference class.

SAM is a very useful tool for vegetation mapping and have included it as part of the "standard" image processing package for the project data. The method works well if the vegetation classes are well separated spectrally, but not as well with less distinct classes.

Lesson 5: Spectral Angle Mapper

CD-ROM: *SERDP Training & Data CD*

Required Files: `Ykm1a_reflectance.img`
                               `Ykm1a_MNF.img`
                               `endmembers_reflectance.sli`
                               `endmembers_MNF.sli`

Generated Files: `Ykm1a_ref_SAM.img`
                               `Ykm1a_ref_SAM_rule.img`
                               `Ykm1a_MNF_SAM.img`
                               `Ykm1a_MNF_SAM_rule.img`

The SAM algorithm implemented in ENVI takes as input a number of training classes or reference spectra from ASCII files, ROIs, or spectral libraries. It calculates the angular distance between each spectrum in the image and the reference spectra or endmembers in *n*-dimensions. The result is a classification image showing the best SAM match at each pixel and a rule image for each endmember showing the actual angular distance in radians between each spectrum in the image and the reference spectrum. Darker pixels in the rule images represent smaller spectral angles and, thus, spectra that are more similar to the reference spectrum. The rule images can be used for subsequent classifications with different thresholds to decide which pixels are included in the SAM classification image.

Execute Spectral Angle Mapper (SAM)

1. Open the reflectance data file `Ykm1a_reflectance.img`. In Available Bands list, right click on `Ykm1a_reflectance.img` to load a true color image into a new display. This image is based on pre-selected bands.

2. In ENVI's main menu, select **Classification → Supervised → Spectral Angle Mapper** to start the SAM endmember selection process.

3. In the **Classification Input File** dialog, select `Ykm1a_reflectance.img`.

4. Click **Spectral Subset** to see which bands have been masked out. Click **OK** when ready to proceed.

5. In the **Endmember Collection: SAM** dialog, select **Import → From Spectral Library**.

6. In the **Spectral Library Input File** dialogue, choose `endmembers_reflectance.sli`. Note: If file isn't listed, click **Open File**.

7. In the **Input Spectral Library** dialogue, select all the endmembers.
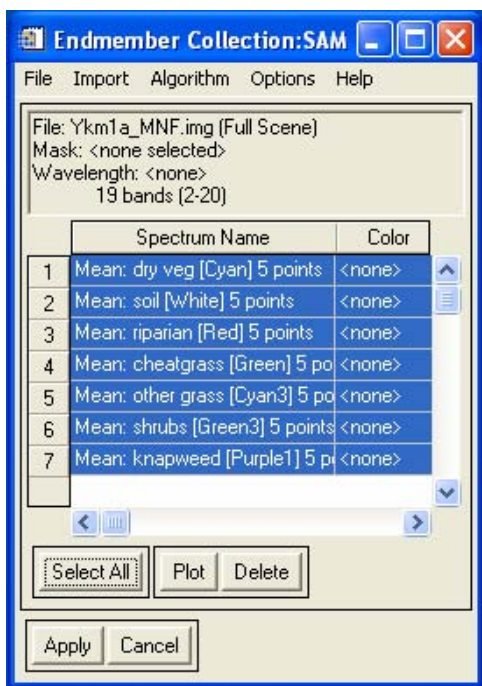

Figure 5-8: Endmember collection for Yakima region.

8. Click **OK** to load all of the endmember spectra into the **Endmember Collection:SAM** dialog.

9. In the **Endmember Collection:SAM** dialog menu, select **Options** → **Plot Endmembers**.

10. **Endmember Spectra** → **Options** → **Stack Plots** for improved comparison of spectral features. Right-click in the plot window, select **Plot Key** from the shortcut menu to display the legend. Enlarge the window so you can see the spectra better.

11. In the **Endmember Collection:SAM** dialog, click **Apply**. Change the **Set Maximum** Angle to None. Enter output file names in the **Spectral Angle Mapper Parameters** dialog, and click OK.

12. Repeat steps 5 through 15 with `Ykm1a_MNF.img`. Use MNF bands 2-20.

Review SAM Results

1. **Select File →Open Image File →** `Ykm1a_ref_SAM.img` to open the SAM classification image. The classification image will appear in the **Available Bands List dialog.**

2. The classified image (`Ykm1a_ref_SAM.img`) consists of one band with coded values for each endmember. For example, hyacinth is an endmember.

3. In the **Available Bands List** dialog, select **Gray Scale**.

4. Load `Ykm1a_ref_SAM.img` into Display #2.

5. Use **Tools → Color Mapping → Class Color Mapping** to identify the corresponding endmember colors.

6. In Display #2 use **Cursor Location/Value** to identify endmember values.

Notice that the number of pixels displayed as a specific endmember is a function of the threshold used to generate the classification. SAM is a similarity measure, not an identifier. Therefore, it is up to the user to determine an appropriate threshold for this method.

Color Codes Used for Endmembers

Category	Color
Dry vegetation	Cyan
Soil	White
Riparian	Red
Cheatgrass	Green
Other grass	Cyan3
Shrubs	Green3
Knapweed	Purple

7. **File → Open Image File** and select `Ykm1a_ref_SAM_rule.img`.

The rule image has one band for each endmember. The pixel values represent the spectral angle in radians. Lower spectral angles (darker pixels) represent better spectral matches to the endmember spectrum. In the Available Bands List dialog, one band for each endmember will appear.

8. In the **Available Bands List** dialog, select **Gray Scale → Display → New Display**, select a band, then click **Load Band** to Display #3.

9. To link all three displays: **Tools → Link → Link Displays → OK**. Evaluate all three images using the **Cursor Location/Value** and the **Z Profiler**.

10. In Display #3, select **Tools → Color Mapping → ENVI Color Tables**.

11. Use **Stretch Bottom** and **Stretch Top** to adjust the SAM rule thresholds. Highlight those pixels with the greatest similarity to the selected endmember.

12. Pull **Stretch Bottom** all the way to the right and **Stretch Top** all the way to the left to highlight the most similar pixels in white.

13. Move **Stretch Bottom** to the left to show only the best SAM matches in white (and reduce the number of highlighted pixels).

14. Use **Rule Image** color composites or **Image Animation** to compare individual rule images. Notice that pure white pixels are the unclassified areas and do not change.

15. Repeat the process with each SAM Rule Image. **Select File → Cancel** when finished to close the **ENVI Color Tables** dialog.

16. Repeat steps 1 through 15 with `Ykm1a_MNF_SAM.img` and `Ykm1a_MNF_SAM_rule.img`. Note: The MNF image is the result of a data reduction process which will be discussed in a later section.
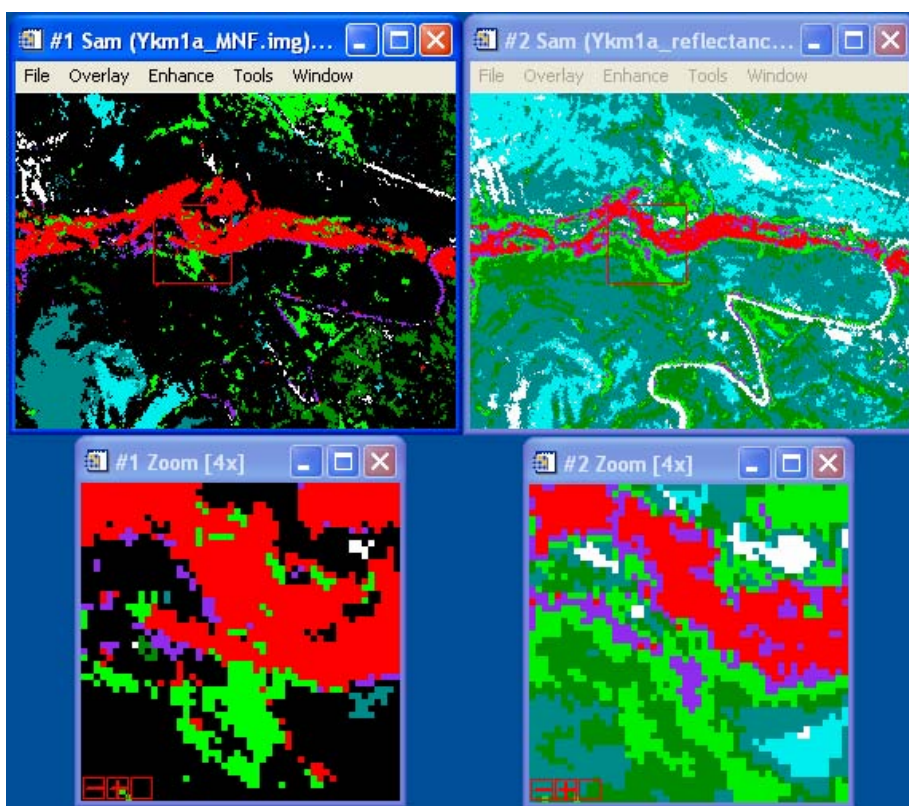

Figure 5-9: Spectral Angle Mapper results for MNF image (left) and reflectance image (right)

Practice Questions (Optional)

• What ambiguities exist in the SAM classification when you work with the images and spectra in this section?

• What factors could affect how well SAM matches endmember spectra?

• Can you see the topographic shading effects in the SAM data? Why or why not?

• How could you select better endmembers to correct some of the ambiguities in the SAM classification?

5.2.2   Maximum Likelihood Classification

The maximum likelihood classifier evaluates the variance and covariance of the spectral clusters defined by the user and uses this information to develop a probability density function surface as shown in Figure 5-10.
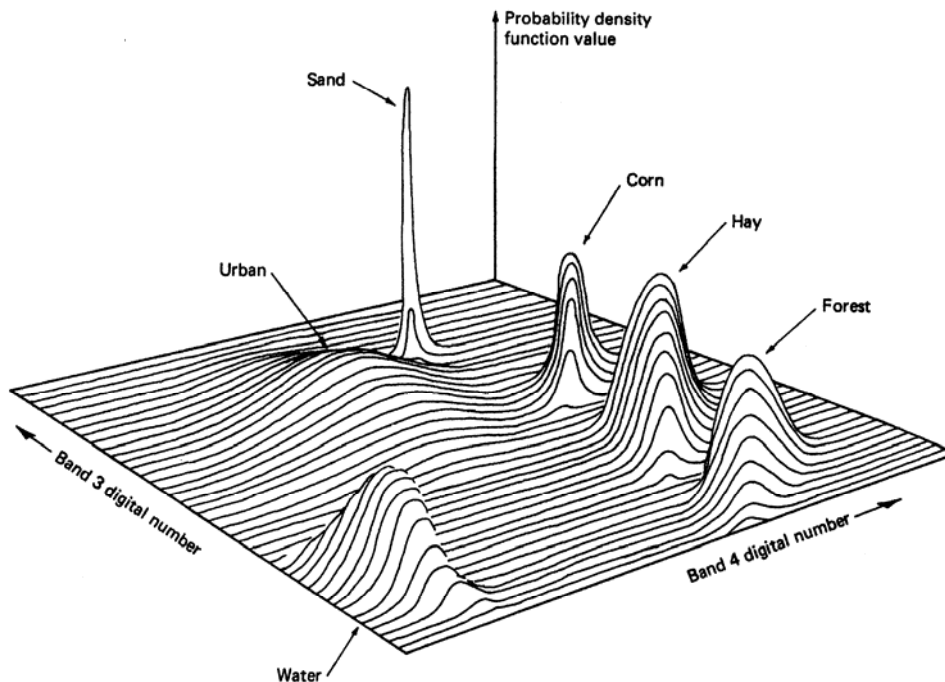
Figure 5-10: Probability density function for pixels in a hypothetical image using the Maximum Likelihood classifier (Lillesand and Kiefer, 2000).

For each pixel, the probability of its belonging to each class is evaluated, and then it is assigned to the class of the highest probability. This technique can be furthered fine-tuned by applying weighting factors to the probability functions for each class based on user knowledge of *a priori* probability and the goals of the classification.

For instance, if based on a field ecological study, and the user knows that there is a low probability of pixels with a sandy soil containing a particular type of vegetation, then the category can be weighted lightly in the sandy soils areas (as defined by the GIS). This lowers the probability of the pixel being classified as the unlikely type of vegetation and pixels that are in borderline areas of the cluster will be assigned to the next closest cluster.

Lesson 6: Maximum Likelihood Classification

CD-ROM: *SERDP Training & Data CD*

Required Files: `Ykm1a_MNF.img`
                        `endmembers.roi`

Generated Files: `Ykm1a_MNF_maxlike.img`

Use **Supervised** Classification to cluster pixels in a data set into categories corresponding to user-defined training classes.

Training classes are groups of pixels (ROIs) or individual spectra. Select them as representative areas or materials that you want mapped in the output. Try to select ROIs that are homogenous.

Prior to performing supervised classification, define the training classes in one of two ways: using the **Endmember Collection** dialog to select spectra or by defining regions of interest (ROIs). The training sites can be defined as multiple irregular polygons, vectors, and/or individual pixels.

Selecting Regions of Interest (ROIs)

To select regions of interest as training classes, click on the desired ROI names in the **Select Classes from Regions** list of available ROIs. Maximum Likelihood requires at least two distinct ROIs.

Applying Maximum Likelihood Classification

Maximum likelihood classification uses the assumption that the statistics for each class in each band are normally distributed. It calculates the probability that a given pixel belongs to a specific class. Unless you select a probability threshold, all pixels are classified. Each pixel is assigned to the class that has the highest probability (i.e., the maximum likelihood).

1. In ENVI's main menu **File** → **Open Image File** → `Ykm1a_MNF.img`.

2. In the **Available Bands List** dialog → **RGB Color** and select the first three bands- 1, 2 & 3. Click **Load RGB**.

3. **Basic Tools** → **Regions of Interest** → **ROI Tool**.

4. Select **File** → **Restore ROIs** → `endmembers.roi`. Minimize ROI tool.

5. Select **Classification** → **Supervised** → **Maximum Likelihood**.

6. In the **Classification Input File** dialog, select the input file `Ykm1a_MNF.img` by clicking once.

7. Select **Spectral Subset** Bands 2 to 5.

8. Click **OK**. The **Maximum Likelihood Parameters** dialog appears.

9. Select all items from the list of regions.

10. Set the probability threshold to **None**.

11. Enter the default value of 1 as the **Data Scale Factor**. The scale factor is a division factor used to convert integer scaled reflectance or radiance data into floating point values. For example, for reflectance data scaled into the range of zero to 10,000, set the scale factor to 10,000.

12. Specify `Ykn1a_MNF_maxlike.img` as the name for the output.

13. In **Output Rule Images** select **No**.

14. Click **OK**.

15. After process is complete, click **New** Display and select `Ykm1a_MNF_maxlike.img` from **Available Bands** list. Click **Load Band**.

6   Data Reduction

Because of the enormous volume of data contained in a hyperspectral data set, data reduction techniques are an important aspect of hyperspectral data analysis. Reducing the volume of data, while maintaining the information content, is the goal of the data reduction techniques covered in this section.

The images created by the data reduction techniques are then used as inputs to the classification techniques described in Section 5.2.

6.1   Masking

Masking reduces the spatial extent of the analysis by masking out areas of the image which do not contain data of interest. Two common masks used in this project were hydrology and road masks.

The hydrology masks used a buffer from the hydrology GIS layer to mask out non-riparian zones which were analyzed for riparian invasive plants.

The road masks used a combination of the roads layer and a vegetation index transform of the image data to find non-vegetated areas. These areas were then masked out to remove these areas from the vegetation classification.

Prior to the supervised classification a mask of the waterways was digitized for each flight line. This mask was then used during the classification process to restrict the analysis to areas of interest and to further reduce the amount of data processed in each scene

During the masking process, the user compares the mask carefully to the original image to verify that only non-essential data is removed. If there is any doubt whether data is important, it is left in the data set.

6.2 Spectral Indices

Another important data reduction technique that can reduce the number of bands used in the analysis but retain the most important information is to create spectral indices relevant to plant properties and functions.

Spectral indices, in their simplest incarnation, are ratios or linear combinations of two or more spectral bands (Jackson & Huete 1991), typically at one wavelength affected by the absorber of interest and one reference wavelength. As shown in Figure 6-1, leaf reflectance spectra are dominated by the absorbance of chlorophyll and water. Not surprisingly, spectral indices are frequently derived to estimate these parameters. The most common are vegetation indices, especially the normalized difference vegetation index (NDVI), which has been used in numerous published studies (> 1000 found searching ISI WebOfScience). These indices are sensitive to vegetation cover and provide an estimate of the "greenness" of an area. As such, they have shown to be correlates of percent cover of vegetation, leaf area index (LAI), and plant biomass.
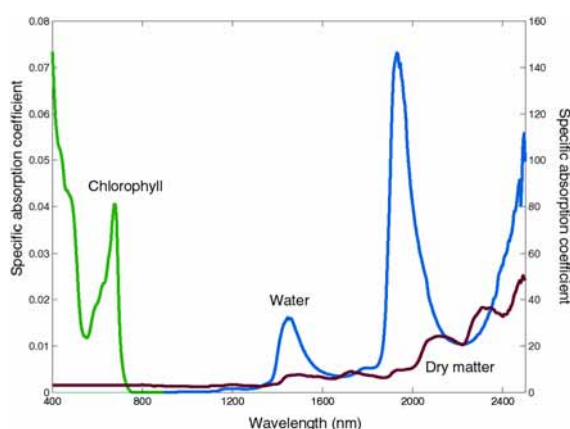

Figure 6-1: Specific absorption coefficient of chlorophyll a+b (cm^2 μg^{-1}) on the left axis, of water (cm^{-1}) and dry matter (cm^2 g^{-1}) on the right axis (from Ustin & Jacquemoud 2002).

Hyperspectral instruments allow the elaboration of spectral indices beyond simple greenness indicators. Since AVIRIS samples 126 bands between 400 and 2500 nm, it allows the detection of narrow absorption features. High spectral resolution instruments have used these narrow features to elucidate foliar chemistry such as lignin, starch, protein and cellulose concentrations (Card et al. 1988, Fourty & Baret 1998, Curran 1989). These techniques have proven so successful in laboratory settings that they can replace traditional wet chemistry extraction methods (Curran 1989). The detectability of foliar chemistry suggests a potential for developing spectral indices sensitive to subtle variations in leaf physiology. A plethora of physiological indices now exist for detecting a range of leaf properties.

Physiological index based classifications allow for explicit interpretations of physical properties and therefore, can be preferable to typical statistical classifications for which the basis of the differences in an image data set are unknown. Statistical classifications integrate all information contained within class spectra. They capitalize on the fact that classes are different, but do not provide any information regarding the basis of that differentiation. Furthermore, these differences may be habitat variables that only indirectly relate to vegetation cover. Physiological indices do not contain this limitation because they estimate physical parameters which may then be used to differentiate between classes. For example, Underwood et al. (2003) successfully mapped infestations of the invasive iceplant at Vandenberg AFB using a water band index (Ustin et al., 1998), indicating that physiological differences (in this case succulence contributing to foliar water content) contribute to spectral uniqueness.

The utility of physiological indices for mapping of invasive plant species was investigated at all seven 7 bases spanning a variety of physical and ecological conditions. Of particular interest is the generality of indices across sites. Do the same collection of indices function well across ecological conditions, or are sets of indices site- or species-specific? Nineteen (19) physiological indices of three basic types: pigment indices, foliar water indices, and foliar chemistry indices were used (Tables 6-1).

Pigment indices focus on the absorptions of photosynthetic pigments. A number specifically highlight chlorophyll, the predominant photosynthetic pigment of green vegetation. In addition to chlorophyll, however, plants possess a variety of accessory pigments; examples include carotenoids and xanthophylls. Relative concentrations of these accessory pigments may vary between species or between individuals of the same species experiencing different stress levels (Zarco-Tejada 1998). See Table 6-1 for a listing of pigment indices studied including descriptions of their sensitivities.

Water absorbs strongly throughout the solar infrared wavelengths. Foliar water content thus heavily influences reflectance throughout these regions (Carter 1991). In particular, vegetation spectra exhibit strong water absorption features at 970 nm, and 1.24, 1.4, and 1.9 μm. The depth of these absorption features corresponds to foliar water content. Invasive plant species with different water uptake (e.g, tamarisk) or storage (e.g, iceplant) strategies than native vegetation are expected to display water index values that reflect these strategies, and thus be readily detectable. An example of a water index is shown in Figure 6-2, below.

Table 6-1: Physiological indices used in vegetation mapping.

Index	formula	details	citation
Pigment indices			
SR, Simple Ratio	$\dfrac{R_{NIR}}{R_R}$	Index of green vegetation cover. Various wavelengths used, depending on sensor.(eg: NIR=845nm, R=665nm)	
NDVI, Normalized Difference Vegetation Index	$\dfrac{R_{NIR} - R_R}{R_{NIR} + R_R}$	Index of green vegetation cover. Various wavelengths used, depending on sensor.(eg: NIR=845nm, R=665nm)	
mNDVI, modified NDVI	$\dfrac{R_{750} - R_{705}}{R_{750} + R_{705}}$	leaf chlorophyll content	Fuentes et al. (2001)
Summed green reflectance	$\sum\limits_{n=500}^{599} R_n$	Index of green vegetation cover.	Fuentes et al. (2001)
PRI, Photochemical Reflectance Index	$\dfrac{R_{531} - R_{570}}{R_{531} + R_{570}}$	Xanthophyll response to light ~ photosynthetic efficiency. Also sensitive to carotenoid/chlorophyll ratio	Rahman et al. (2001)
Red/Green ratio	$\dfrac{R_{600-699}}{R_{500-599}}$	anthocyanins/chlorophyll	Fuentes et al. (2001)
NPCI, Normalized Pigments Chlorophyll Ratio Index	$\dfrac{R_{680} - R_{430}}{R_{680} + R_{430}}$	total pigments/chlorophyll	Peñuelas et al. (1995)
SRPI, Simple Ratio Pigment Index	$\dfrac{R_{430}}{R_{680}}$	carotenoid/chlorophyll *a* content	Zarco-Tejada (1998)
NPQI, Normalized Phaeophytinization Index	$\dfrac{R_{415} - R_{435}}{R_{415} + R_{435}}$	chlorophyll degradation, detects stress at early states	Zarco-Tejada (1998)
SIPI, Structure Intensive Pigment Index	$\dfrac{R_{800} - R_{445}}{R_{800} - R_{680}}$	carotenoid/chlorophyll *a* concentrations	Zarco-Tejada (1998)
PI1, Pigment Index 1	$\dfrac{R_{695}}{R_{420}}$	plant stress status	Zarco-Tejada (1998)
PI2, Pigment Index 2	$\dfrac{R_{695}}{R_{760}}$	plant stress status	Zarco-Tejada (1998)
PI3, Pigment Index 3	$\dfrac{R_{440}}{R_{690}}$	vegetation health, based on chlorophyll fluorescence ratios	Lichtenthaler et al. (1996)
PI4, Pigment Index 4	$\dfrac{R_{440}}{R_{740}}$	vegetation health, based on chlorophyll fluorescence ratios	Lichtenthaler et al. (1996)

Table 6-2: Physiological indices used in vegetation mapping (continued).

Index	formula	details	citation
Water indices			
NDWI, Normalized Difference Water Index	$\dfrac{R_{860} - R_{1240}}{R_{860} + R_{1240}}$	leaf water content	Gao (1996)
WBI, Water Band Index	$\dfrac{R_{900}}{R_{970}}$	leaf water content	Peñuelas et al. (1997)
Foliar chemistry indices			
NDNI, Normalized Difference Nitrogen Index	$\dfrac{\log\left(R_{1680}/R_{1510}\right)}{\log\left(1/R_{1680}R_{1510}\right)}$	foliar nitrogen concentration	Serrano et al. (2002)
NDLI, Normalized Difference Lignin Index	$\dfrac{\log\left(R_{1680}/R_{1754}\right)}{\log\left(1/R_{1680}R_{1754}\right)}$	foliar lignin concentration	Serrano et al. (2002)
CAI, Cellulose Absorption Index	0.5 * (R$_{2020}$+R$_{2220}$)-R$_{2100}$	based upon cellulose & lignin absorption features, used to discriminate plant litter from soils	Nagler et al. (2000)

Figure 6-2: The water band index (WBI) for vegetation at VAFB (left) and Yuma Proving Ground (right).

Finally, several indices have been proposed to estimate leaf chemistry. As with pigment and water indices, it is probable that different species display distinct proportions of foliar chemicals. For example, the invasive vine kudzu is leguminous, forming nitrogen fixing

symbioses. As a result, it may have higher foliar nitrogen content than native species, and consequently higher normalized difference nitrogen index (NDNI) values. Note that the cellulose absorbance index (CAI; Nagler et al. 2000), in contrast to the other indices studied, estimates the cellulose content of senescent plant matter, emphasizing the distinctions between soil and plant litter. This index may be useful in arid and semi-arid regions where healthy green vegetation is (seasonally) sparse or fall senescence in mixed evergreen and deciduous ecosystems.

All of the parameters estimated by the physiological indices described here are in leaves, and as a result, all tend to be correlated. Obviously the several indices estimating chlorophyll content will be related. Additionally, there are correlations between various physiological parameters within plants, leading to further correlations between physiological indices. For example, the water indices and NDNI are generally highly correlated to NDVI. This is because the more water and nitrogen a plant has access to, resulting in higher leaf water and nitrogen content, the healthier it is and the more it is able to invest in photosynthetic machinery, resulting in higher NDVI. Of interest are the slight differences in spectral leaf traits across species and environments.

The physiological indices images are created by deriving a variance/covariance matrix, eliminating the most highly correlated bands, and then "stacking" the single band indices into a final multi-band image. These indices images can be used to encapsulate the information contained within vegetation spectra in meaningful and interpretable ways. Our use of physiological indices will provide several benefits. Of immediate concern, they are a valuable tool that will facilitate the detection and mapping of invasive plant species. Furthermore, our investigation of their portability across species and sites will enhance understanding of the properties of physiological indices.

Lesson 7: Band Math & Physiological Indices

CD-ROM: *SERDP Training & Data CD*

Required Files: Ykm2a_reflectance.img

Generated Files: Ykm2a_ind01_SR.img
Ykm2a_ind02_NDVI.img
* * *
Ykm2a_ind13_NDNI.img
Ykm2a_index_stack.img
Ykm2a_NaN_mask.img
Ykm2a_index_stack.txt
Ykm2a_index_stack.sta
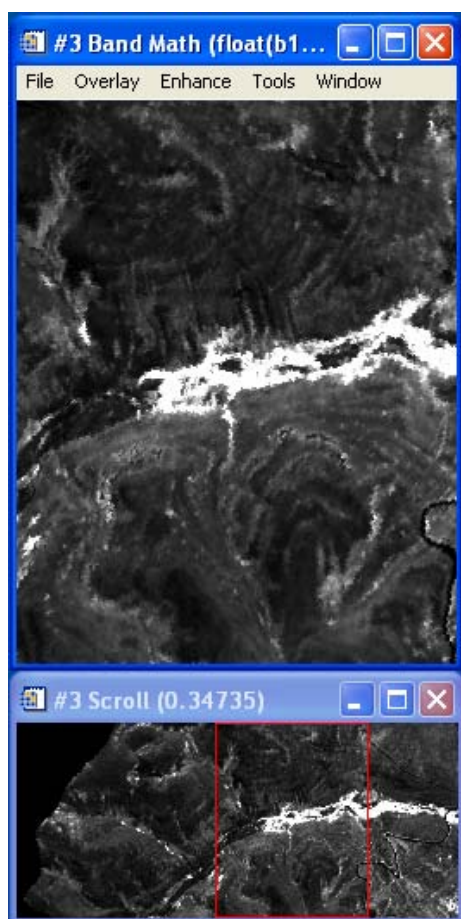
Band Math Exercise

1. Calculate vegetation indices, covariance statistics, and correlation matrices with ENVIs Band Math function.

2. The first vegetation index, (shown in Figure 6-3) the Simple Ratio Index (SR), is a measure of photosynthetic efficiency that is calculated in these steps.

3. In ENVI's main menu, select **File → Open Image File →** Ykm2a_reflectance.img

4. In the **Available Bands** list, right click on Ykm2a_reflectance.img to "**Load True Color to New Display**".

5. Select **Basic Tools → Band Math**.

6. In the **Band Math Expression** dialog, enter float(b1)/float(b2).

7. Click **OK.**

Figure 6-3: Water Band Index (WBI) calculated for a fragment of Yakima region.

8. In the **Variables to Bands Pairing** dialog, click on **B1** (undefined).

9. Select Band 6 (845nm) for B1.

10. Click on **B2** (undefined).

11. Select Band 9 (665 nm).

12. Name the output `Ykm2a_ind01_SR.img`

13. Click **OK**.

14. Repeat steps 4 - 13 for some of the other vegetation indices described in Table 6-1 such as those listed here.

Note: Some bands were noisy and were removed during the preprocessing stage. The wavelengths corresponding to these bands will not be available.

- NPCI, Normalized Pigments Chlorophyll Ratio Index: float(b1-b2)/float(b1+b2). Click on the previous **Band Math** expression and assign new variable values.

- SRPI, Simple Ratio Pigment Index: float(b1)/float(b2).

- WBI, Water Band Index: float(b1-b2)/float(b1+b2). Here you can also click on the previous **Band Math** expression and assign different variable values (see picture to the left).

- mNDVI, modified NDVI.

- SIPI, Structure Intensive Pigment Index: float(b1-b2)/float(b1-b3)

- NDNI, Normalized Difference Nitrogen Index: alog10(float(b1)/float(b2))/alog10(1/(float(b1)*float(b2)))

- Other indices of your choosing.

Combine Bands into a Single Image

Combine all bands into a single image to calculate covariance statistics and a correlation matrix. Use ENVI's layer stacking function for georegistered images.

1. In ENVI's main menu, select **Files → Save Image As → ENVI Standard**.

2. Click **Import** and select vegetation index files you've just created: hold the **Ctrl** button to select multiple files (`Ykm2a_ind01_SR.img`, `Ykm2a_ind02_NDVI.img,…,Ykm2a_ind13_NDNI.img`).

3. Name the file `Ykm2a_index_stack.img`

4. Click **OK**.

5. In ENVI's main menu, select **File → Edit ENVI Header → ENVI Standard**.

6. Select `Ykm2a_index_stack.img`.

7. Click **Edit Attributes** and select **Band Names**. Change band names to the appropriate index names – if needed.

8. Click the **Display** button and select **New Display**. Click **Load RGB** to load the **New Stacked Layer** into a new display.

Checking for Independence of Calculated Indices

Calculating a correlation matrix is a fairly simple method for verifying the usefulness of these indices. Mask out zero and –NaN values to ensure that useful values are generated.

Build the Mask

1. Load `Ykm2a_index_stack.img` into Display 1.

2. Select **Basic Tools → Masking → Build Mask**.

3. In the **Mask Definition dialog**, select **Options → Import Data Range**; enter "**0**" for **Data Min Value** and **Data Max Value**, and check **Mask pixel if ANY band matches range.**

4. In the **Mask Definition** dialog, select **Options → Mask NaN.**

5. Select **Options → Selected Areas Off**.

6. In the **NaN Options** dialog, check **Mask pixel if ANY band matches NaN** option.

7. Click **OK**.

8. Name the file `Ykm2a_NaN_mask.img`, click **Apply**.

9. In the **Mask Definition** dialog, click **Cancel**.

Calculate Statistics and Covariance Image

1. In ENVI's main window, select **Basic Tools → Statistics → Compute Statistics**.

2. Select `Ykm2a_index_stack.img`.

3. In the Compute Statistics Input File dialog, click **Select Mask Band** and select mask you've just created: `Ykm2a_NaN_mask.img`.

4. Click **OK**.

5. Check **Covariance, Output to a Statistics File**, and **Output to a Text Report File**.

6. Enter `Ykm2a_index_stack.sta` for the statistics file and `Ykm2a_index_stack.txt` for the text report file.

7. Click **OK**.

8. Maximize the **File Statistics Report** and interpret the correlation matrix. A high absolute value (close to 1 or negative 1) indicates that the two bands are highly correlated.

Frequently Asked Questions & Troubleshooting Guide for Calculating Physiological Indices and Performing Classifications

My index image is all zeros.

All zeros occur because the raw image is saved as an integer data type. Unless told otherwise, the computer will set the type of the output band math image to integer as well. This results in the truncation of decimal places. An index that ranges between -1 and 1, which many of them are normalized to do, will be saved as 0. When entering your band math expressions, you must convert the input bands to decimals (or floating point values, in computer speak). Example calculations are:

-   NDVI = float(b1-b2)/float(b1+b2)

-   SR = float(b1)/float(b2)

-   NDNI = alog10(float(b1)/float(b2))/alog10(1/(float(b1)*float(b2))

How do I combine multiple indices into a single image?

If your images are georegistered, combining multiple indices can be accomplished in ENVI using **Layer Stacking**:

1. Go to **Basic Tools** → **Layer Stacking**.

2. Select and order bands as desired. The bands you are stacking must be georegistered for this process. If your indices are not georegistered, you can combine bands by selecting **File → Save File As → ENVI Standard**. From here you can select and reorder bands.

   After you have a consolidated index image, rename the bands to be more meaningful as follows:

3. In the **Available Bands** window, right-click on the image name. Select **Edit Header → Edit Attributes → Band Names**. You can now rename each band to correspond with the index.

4. Click OK when you are finished.

How do I decide which indices to use for a classification?

Many of these indices are correlated with each other. Multiple highly correlated bands will provide no new information to be used by a classifier algorithm. It may even be detrimental to use highly correlated bands due to the noise contained within each band.

1. To determine correlations go to **Basic Tools → Statistics → Compute Statistics**.

2. Select your physiological index image, mask out margin areas.

3. Click **OK**.

4. Select **Calculate Covariance Statistics**. Choose a text report rather than a covariance image.

5. Click **OK**.

The output will include summary statistics (mean, standard deviation) of each band, a covariance table, and a correlation table. The correlation table is of interest. It displays the correlation between all pairwise comparisons of bands (indices). As you can see, this matrix is diagonally symmetric. A high absolute value (close to 1 or negative 1) indicates that the two bands are highly correlated. When performing a classification, only one index from a group of highly correlated ones should be used.

Noisy indices also should be avoided as inputs to a classification. You can determine noise by visual inspection. Open each index image as a greyscale, and check how grainy or speckled it is.

The correlation matrix I've calculated has a lot of zeros, -NaN, or infinite values.

Lots of zeros, -NaN and infinite values come from indices being undefined at some pixels. For example, ratio indices will be undefined in the margin area since they contain

a zero in the denominator. To avoid this problem, when you calculate your correlation matrix, you should mask out all undefined areas.

1. Go to **Basic Tools** → **Masking** → **Build Mask**.

2. In the **Mask Definition Window**, go to **Options** → **Mask NaN** values.

3. Repeat for all bands, or check **Mask pixel if ANY band matches NaN** option in the **NaN Option** dialog.

4. To mask out these areas, select **Options** → **Selected Areas Off**.

5. Type a file name, and click **Apply**.

6.3   Continuum Removal

Continuum removal reduces the spectral extent of the data by focusing the analysis on bands within specific absorption features. It is a process which compares absorption features of reference and image spectra. The continuum is a convex hull fit over the top of a reflectance spectrum using straight line segments to connect local spectra maxima, as shown in Figure 6-4. The continuum line segments provide a common baseline from which to compare individual absorption features.

The continuum reflectance, R_c, is removed by dividing it into the actual reflectance, R, for each pixel in the image.

$$r' = \frac{R}{R_c}$$

The resulting image spectra, r', are equal to 1.0 where the continuum and the spectra match, and less than 1.0 where absorption features occur as shown in Figure 6-5.
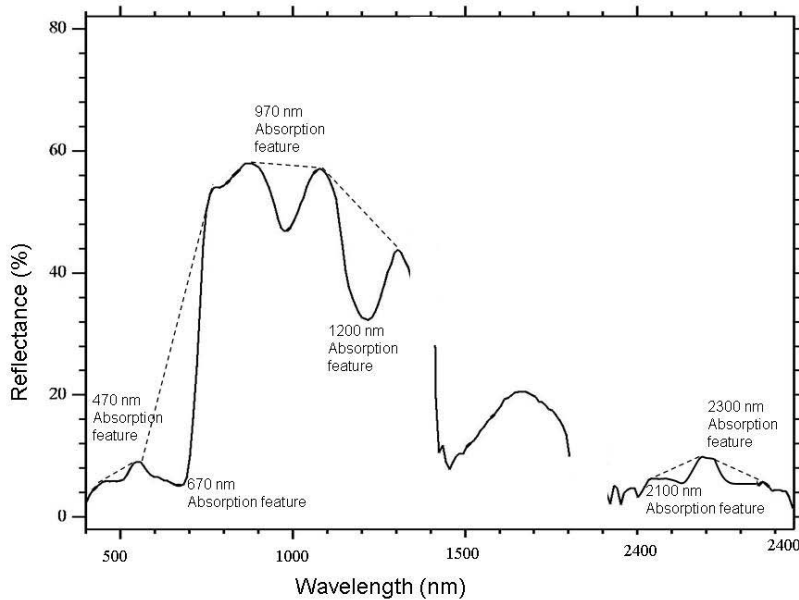
Figure 6-4: Example spectrum showing common absorption features and connecting line segments.
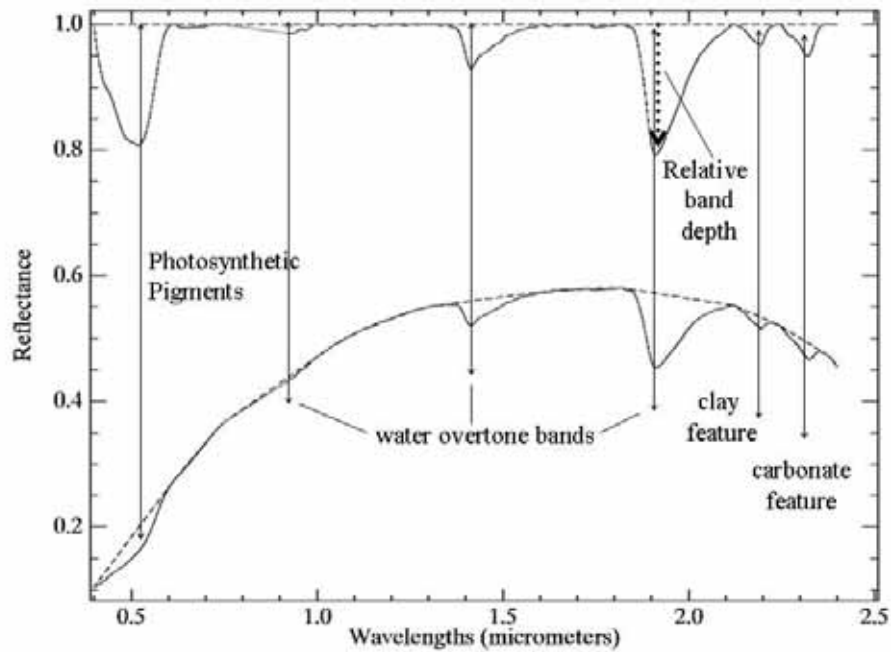
Figure 6-5: Example of a continuum removal calculation typical plant and soil absorption features.

An absorption feature at 2100 nm is shown in the reflectance plot in Figure 6-5 with a continuum line connecting the shoulders of the absorption feature. The upper panel shows the same feature rescaled as a fractional % of the continuum reflectance.

The continuum removal technique works best when the target invasive plant species have absorption features significantly different than the background pixels. This technique may be particularly appropriate in differentiating between the following classes.

- Succulent vs. non-succulent vegetation
- Green vegetation vs. dead vegetation
- Vegetation vs. soil

The absorption feature is selected after a review of the image spectra and the hyperspectral data is subset to the region containing the absorption features of interest.

Lesson 8: Continuum Removal

CD-ROM: *SERDP Training & Data CD*

Required Files: `Ykm2a_reflectance.img`

Generated Files: `Ykm2a_cont_rem.img`

Use **Continuum Removal** to normalize reflectance spectra to compare individual absorption features from a common baseline. The continuum is a convex hull fit over the top of a spectrum utilizing straight line segments that connect local spectra maxima. The first and last spectral data values are on the hull. Therefore the first and last bands in the output continuum-removed data file are equal to 1.0.

Using different spectral subsets gives different results, so you should spectrally subset the data to the region containing the absorption features of interest.

The continuum is removed by dividing it into the actual spectrum for each pixel in the image. The resulting image spectra are equal to 1.0 where the continuum and the spectra match, and less than 1.0 where absorption features occur. Continuum removal can be performed on data files or on individual spectra in a plot window.

1. In ENVI's main menu select **File → Open Image File** select `Ykm2a_reflectance.img`. Load as true color composite (R: Band 28, G: Band 19, B: Band 10).

2. Select **Spectral → Mapping Methods → Continuum Removal.** Click on `Ykm2a_reflectance.img` once.

3. Spectrally subset the image by turning "off" the bands around the absorption feature at 1200nm. In the **Continuum Removal Input File** dialog select **Spectral Subset.** Scroll up to the absorption band of interest (Band 90, 1201nm). Hold **Ctrl** down while selecting Bands 81-100. This will turn the bands "off". Click **OK**.

4. Enter output file name `Ykm2a_cont_rem.img` in the Continuum Removal Input Parameters dialog and click OK.

5. Click the **Display** button, and add a new display.

6. Click the **RGB** radio button, and select Band 28 (red), Band 19 (green) and Band 10 (blue). Click **Load RGB**.

Compare the two images by linking them and displaying the z-profile of each image.

7. From one of the main image display windows, select **Tools** → **Link Displays**. Click **OK** in the **Link Displays** dialog box.

8. Right click on the main image window of the first image (reflectance). Select **Z-Profile**. This will display the spectral profile of the selected pixel. Repeat for the second image (continuum-removed).

You should now have both images linked and two z-profiles displayed. Notice the difference in the spectrum of the images. Move the zoom box around and display the spectral profiles of different pixels. Notice the absorption feature is now removed.

6.4   Minimum Noise Fraction

This section is partially adapted from the ENVI tutorial manual (Research Systems, 2002)

The minimum noise fraction (MNF) transformation is used to denoise and to reduce the dimensionality of the AVIRIS. The MNF transform as modified from Green et al. (1988) and implemented in ENVI, is essentially two cascaded Principal Components Analysis (PCA) transformations.

The PCA is a transformation method of distilling the data by reducing the data set dimensions. The image data as acquired from the sensor is a data cloud in multi-dimensional space with each band generating an axis of dimension. When the data cloud is viewed in two or three dimensions, the shape of the cloud depends on the bands chosen to supply the axis of the visual plot. Some of the bands are highly correlated and if they are chosen as axes there is very little spread in the data plot. The PCA is a linear transformation that re-orients the axis to capture the information in a non-correlated way. The resultant axes are linear combinations of the previous (band-based) axes.

The PCA does the axis-transformation by performing an eigen decomposition of the sample covariance of the data. The first axis is placed on a trajectory to capture the widest variation in the data cloud. The second axis is orthogonal to the first and captures the second widest variation and so forth. A schematic demonstration of the first two axes is shown in Figure 6-6.
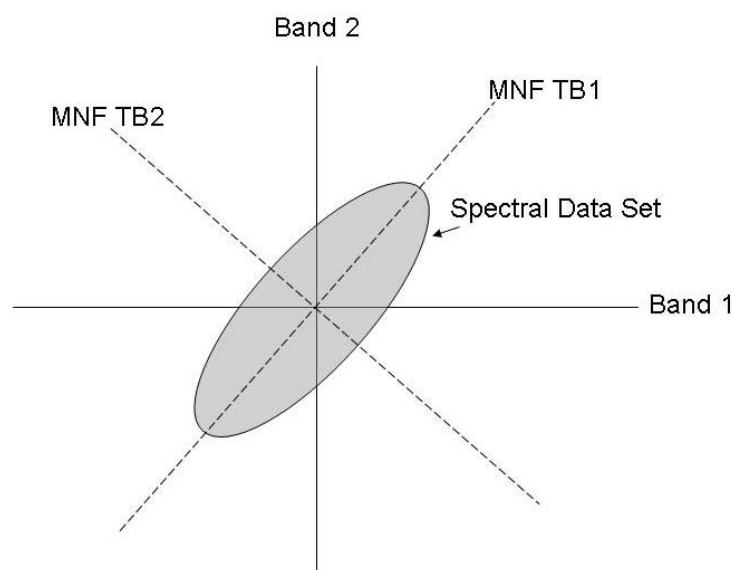

Figure 6-6: Schematic representation of the first two eigenvectors (MNF TB1 and MNF TB2) from a PCA decomposition of a hypothetical data set

The disadvantage to the PCA reduction is that the eigenvectors are based on data variance and not on the signal-to-noise ratio. Since noise can contribute to large data variances, this method is unable to reliably separate signal from noise.

The MNF resolves this problem by using a two-stage PCA. The first transformation, based on an estimated noise covariance matrix, decorrelates and rescales the noise in the data. The second step is a standard PCA transformation of the noise-whitened data. The inherent dimensionality of the data is determined by examination of the final eigenvalues and the associated images.

MNF data space can be divided into two parts: coherent eigenimages and noise-dominated images. By using only the coherent portions, the noise is separated from the data, thus improving spectral processing results and substantially reducing the volume of the image data set.
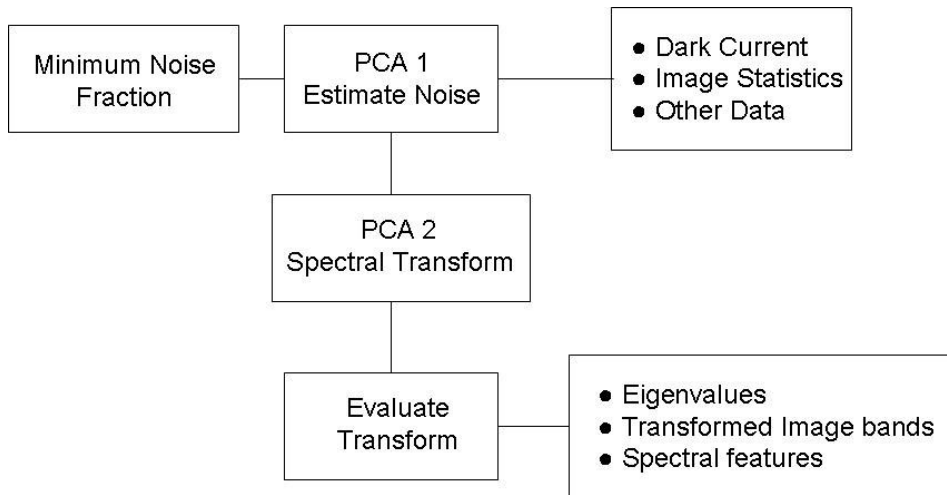

Figure 6-7: MNF Procedures in the ENVI program.

Figure 6-7 summarizes the MNF procedure in ENVI. The noise estimate can come from one of three sources; from the dark current image acquired with the AVIRIS data, from noise statistics calculated from the image, or from statistics saved from a previous transform. Both the eigenvalues and the MNF images (eigenimages) are used to evaluate the dimensionality of the data. Eigenvalues for bands that contain information will be an order of magnitude larger than those that contain only noise. The corresponding images will be spatially coherent when viewed by the user, while the noise images will not contain any discernable spatial information.

Lesson 9: Minimum Noise Fraction

CD-ROM: *SERDP Training & Data CD*

Required Files: `Ykm2b_MNF.img`
                       `Ykm2b_reflectance.img`

Open and Load MNF Image

1. Open the files `Ykm2b_MNF.img` and `Ykm2b_reflectance.img`.

2. In the **Available Bands List** dialog, click on **Gray Scale** and select a band from `Ykm2b_MNF.img`.

3. Click **Load Band**.

Compare MNF Images

4. Select **New Display**. Load several other MNF bands, either one at a time for individual examination or in RGB composites for comparison.

5. Select **New Display**. Right click on `Ykm2b_reflectance.img` to load into display.

6. Use **Z-Profile**, **Image Link**, and **Dynamic Overlay** to compare MNF spectra with apparent reflectance spectra from `Ykm2b_reflectance.img`.

Note: Right click on main image window to bring up **Z-Profile** dialog. Use **Tools →
Link → Link All** Displays to create links and overlay.

7. Determine the relationship(s) between the MNF image and the apparent reflectance images. Relate MNF band number to MNF image quality.
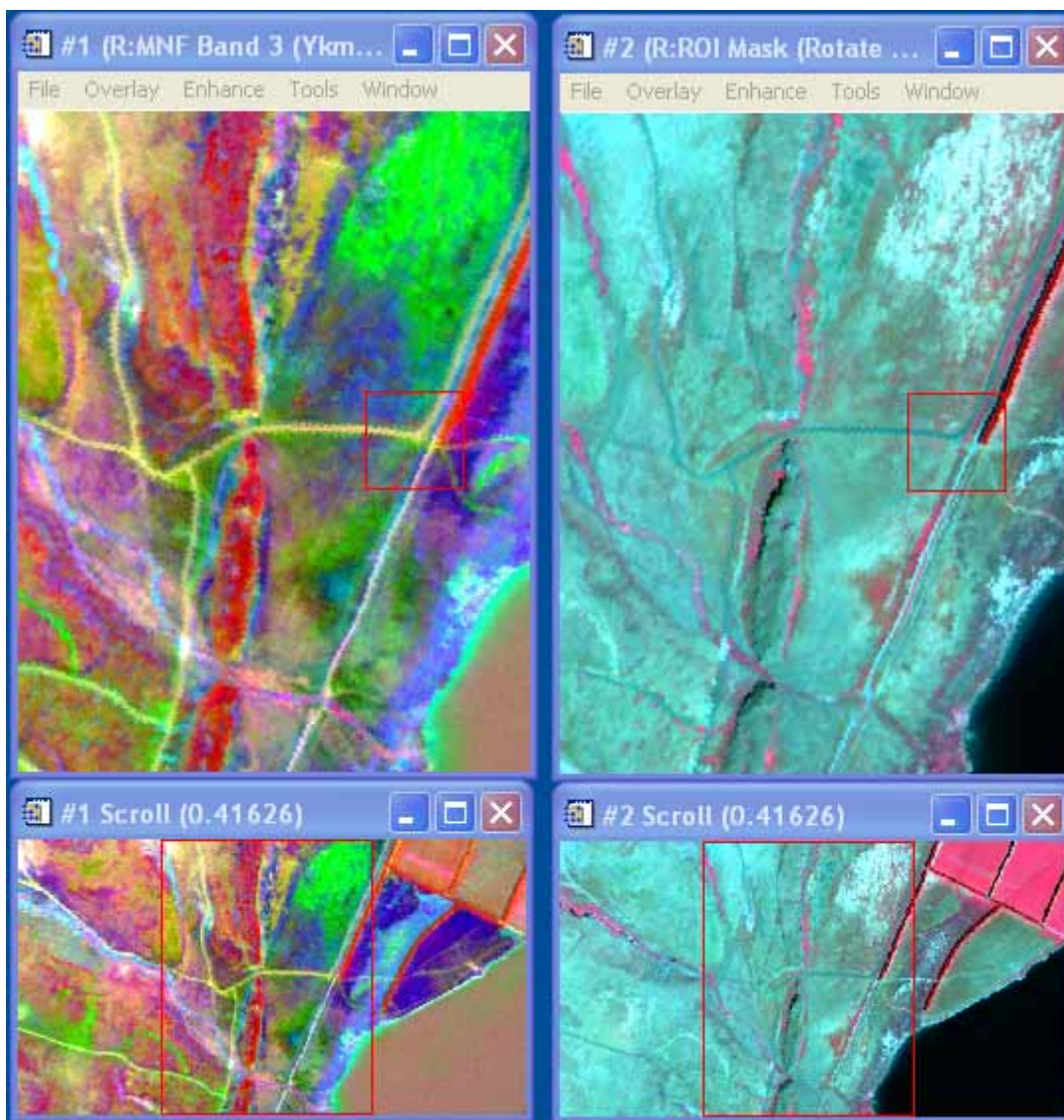
Figure 6-8: Comparison of Minimum Noise Fraction image (left) and reflectance image (right) for Yakima region.

Examine MNF Scatter Plots

8. In the **Main Image** window, use **Tools → 2-D Scatter Plots** for forward and inverse mapping modes.

9. Load a scatter plot of MNF Band 1 (x-axis) and Band 2 (y-axis).

10. Select **Options → Change Bands** in the **Scatter Plot** window to change bands. Notice the corners (pointed edges) on some MNF scatter plots (Figure 6-9).

Examine the high variance (low band number) MNF bands. Also examine at least one scatter plot of a low variance (high band number) MNF band. Can you see a difference?

Use linked windows, overlays, and Z-profiles to understand the reflectance spectra of the MNF corner pixels. Look for areas where the MNF data stops being pointy and begins being fuzzy. Also notice the relationship between scatter plot pixel location and spectral mixing as determined from image color and individual reflectance spectra. How are these patterns explained? How can you exploit them?
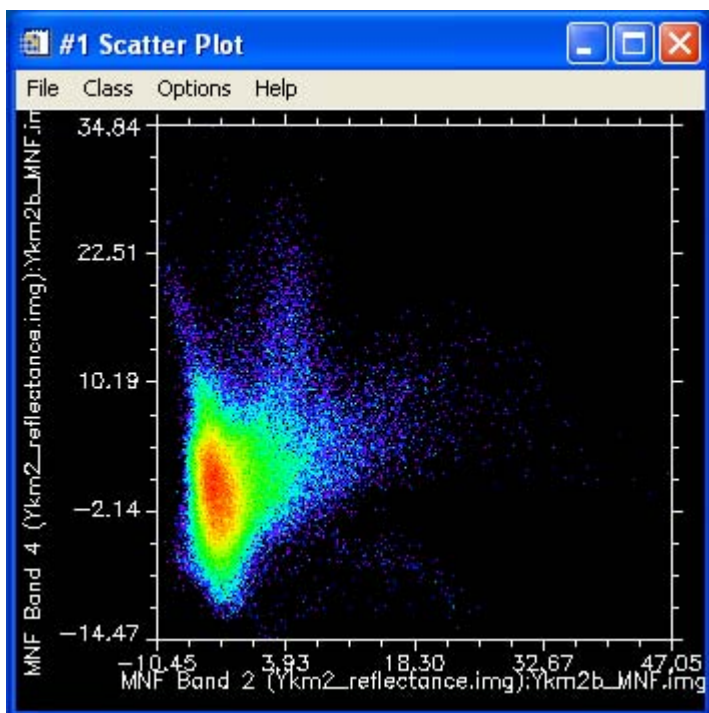

Figure 6-9: MNF 2-D Scatter Plot

Use Scatter Plots to Select Endmembers

1. Derive unmixing endmembers from the data using MNF images and the **2-D Scatter Plot Tool**. Use MNF Bands 1 and 2.

2. In the **Scatter Plot** window, use the ROI drawing functions to circle the extreme few pixels in one or more of the corners or arms of the data cloud.

3. Select desired colors from the **Class** pull-down menu.

4. Use image and scatter plot dancing pixels (double-click and drag with the middle mouse button) to help identify unique areas.

5. Select **Options → Export All** in the **Options** pull-down menu of the scatter plot to export these lists of pixels as ENVI Regions of Interest (ROI).

6. Load these ROIs into the window displaying the apparent reflectance data by choosing selecting **Overlay → Region of Interest** in the **Main Image** window menu bar.

7. Continue to select MNF-corner ROIs using different combinations of the first several MNF bands.

8. Use the **Mean for All Regions** menu item in the **Options** pull-down menu of the ROI Tool to extract the mean apparent reflectance spectra of the ROIs.

9. Use the linked windows and Z-profiles to examine the relations between the MNF and reflectance spectra.

Note: Corner pixels on the scatter plots generally make good endmember estimates. However, note also the occurrence of overlapping or repeat ROIs. This limitation results from examining the data in a pairwise fashion (2-D).

10. Close the 2-D scatter plot by selecting **File → Cancel** in the **Scatter Plot** window.

7   Analyzing at the Sub-Pixel Level

The analyses discussed previously in Section 5, are used to assign pixels to classes, with one class per pixel. However, pixels are rarely pure and each pixel contains a multitude of objects. The previous classification techniques address this by defining mixed classes, where the spectral center of the cluster is the combination of the spectra of the individual objects. The class a pixel is assigned to is dependent on the dominant spectral feature of the pixel and the definition of the classes. The target invasive plant species is classified as "invasive plants" where the pixel is dominated by the invasive plant species, or as an "invasive plant species - native vegetation mixture" when the pixel is dominated by the background vegetation.

This technique works fairly well with invasive plant species that grow in thick clusters and produce a fairly strong signal within the pixel. However, if the invasive plant species of concern is a non-dominant feature in the 4 meter pixels, it may not show up in the pixel-level classification. The next level of mapping techniques assigns pixels to multiple classes and estimates the portion that each class contributes to the overall signal from the pixel.

The following sections discussion spectral mixing models(Research Systems, Inc., 2002) are adapted with permission from the ENVI tutorial discussions (Research Systems Inc., 2003).

7.1   Mixing Models

Spectral mixing occurs when materials with different spectral properties are represented by a single image pixel. The simplest model of a mixed spectrum is a linear one, in which the spectrum is a linear combination of the pure spectra of the materials located in the pixel area, weighted by their fractional abundance. The linear spectral unmixing model assumes no interaction between materials in a pixel. If each solar photon only sees one material, these signals add (a linear process). This is at best an approximation, but this assumption does appear to work well in many circumstances (Boardman and Kruse, 1994).

Mixed pixels can be visualized as points in *n*-dimensional scatter-plot space (spectral space), where *n* is the number of bands. If only two endmembers mix in two dimensions, the mixed pixels will fall along a line (Figure 7-1A). The pure endmembers will fall at the two ends of the mixing line. If three endmembers mix, then the mixed pixels will fall inside a triangle (Figure 7-1B). Higher dimensions of mixing are represented by higher dimension geometric figures (e.g.: four endmembers mix within a tetrahedron, etc.).
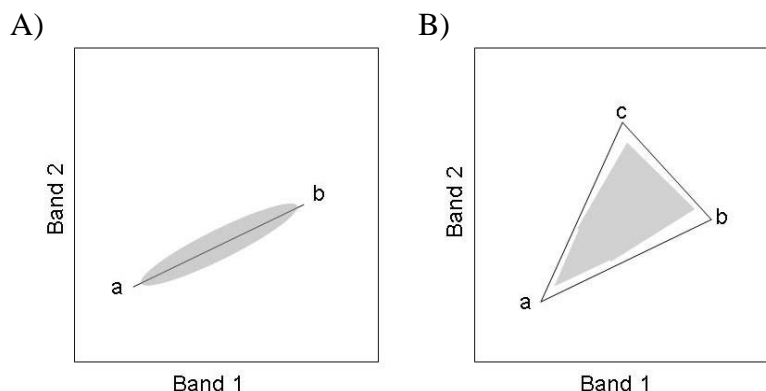
Figure 7-1: Geometric mixing model. A) two-dimensional mixing (two endmembers). B) three-dimensional mixing (three endmembers).

Mixtures of endmembers fill in between the endmembers. All mixed spectra are interior to the pure endmembers, inside the simplex formed by the endmember vertices, because all the abundances are positive and sum to unity.

This convex set of mixed pixels can be used to determine how many endmembers are present and to estimate their spectra. The geometric model is extensible to higher dimensions where the number of mixing endmembers can be up to one more than the inherent dimensionality of the mixed data. Typically however, the number of spectrally unique endmembers is significantly below the inherent dimensionality and on a practical basis, is probably on the order of half that number.

Spectral mixture analysis (SMA) has been one of the most successful techniques for quantifying vegetation abundances (Ustin et al., 1998; Elmore et al., 2000). SMA models the reflectance spectrum of each pixel as a linear combination of the spectral signatures (endmembers) of the dominant ecosystem components (e.g. foliage, soil, rock, senescent and woody vegetation). These fundamental landscape components are useful integrators of biochemical and architectural components of canopies (Smith et al., 1990; Ustin et al., 1993, 1996, 1998). SMA has been applied to the decomposition of a mixed pixel or remotely sensed data for over 30 years (Hortwitz et al., 1971) and has been used to identify dynamic ecosystem changes (Adams et al., 1995; Roberts et al., 1997; Ustin et al., 1996, 1998b). It has also been used to map non-native coastal plants (Underwood et al, 2003) the distribution of wetland species (Zhang et al., 1997; Sanderson et al., 1998), and environmental monitoring (Riano et al., 2002; Elmore et al, 2000).

7.2   Practical Unmixing Methods

Two typical and very different types of unmixing are 1) using known endmembers and 2) using derived endmembers.

In the first type, by using a set of known or assumed spectral endmembers, the apparent fractional abundance of each endmember material can be derived for each pixel. These known endmembers can be calculated from the data, obtained from a library of pure materials, or estimated with models.

The mixing endmember matrix is made up of spectra from the image or a reference library. Unmixing analysis is an over-determined linear least squares problem. The mixing matrix is inverted and multiplied by the observed spectra to get least-squares estimates of the unknown endmember abundance fractions. Constraints can be placed on the solutions to give positive fractions that sum to unity. Shade and shadow are included either implicitly (fractions sum to 1 or less) or explicitly as an endmember (fractions sum to 1).

The second unmixing method uses the imaging spectrometer data to derive the mixing endmembers (Boardman and Kruse, 1994). The inherent dimensionality of the field spectra data is determined using a special orthogonalization procedure related to principal components.

• Derive a linear sub-space (flat) that spans the entire signal in the data.

• Project the data onto this subspace to lower the dimensionality of the unmixing and remove most of the noise.

• Find the convex hull of these projected data.

• Shrink-wrap the data by a simplex of *n*-dimensions, giving estimates of the pure endmembers.

• Derived endmembers must give feasible abundance estimates (positive fractions that sum to unity).

Spectral unmixing is one of the most promising research areas in hyperspectral analysis. Analysis procedures using the convex geometry approach already developed for AVIRIS data have produced quantitative mapping results for a variety of materials (geology, vegetation, oceanography) without *a priori* knowledge. Combination of the unmixing approach with model-based data calibration and expert system identification capabilities could potentially result in an end-to-end quantitative, yet automated, analysis methodology.

7.3 Linear Spectral Unmixing (LSU)

This section is adapted from the ENVI tutorial manual (Research Systems, 2002).

Linear Spectral Unmixing (LSU) is a technique for classifying mixed pixels. LSU is based on the assumption that the reflectance at each pixel is a linear combination of the reflectance of each distinct material (or endmember) present within the pixel. For example, if 25% of a pixel contains material A, 25% of the pixel contains material B, and 50% of the pixel contains material C, the spectrum for that pixel is a weighted average of 0.25 times the spectrum of material A plus 0.25 times the spectrum of material B plus 0.5 times the spectrum of material C. So, given the spectrum of a mixed pixel and a set of endmember spectra, the linear unmixing solves for the abundance values of each endmember for every pixel. The number of endmembers must be less than the number of spectral bands and all of the endmembers in the image must be used. If all endmembers are not known, or if you only want to map a few endmembers, use the Matched Filtering or Mixture Tuned Matched Filtering techniques described in the next section.
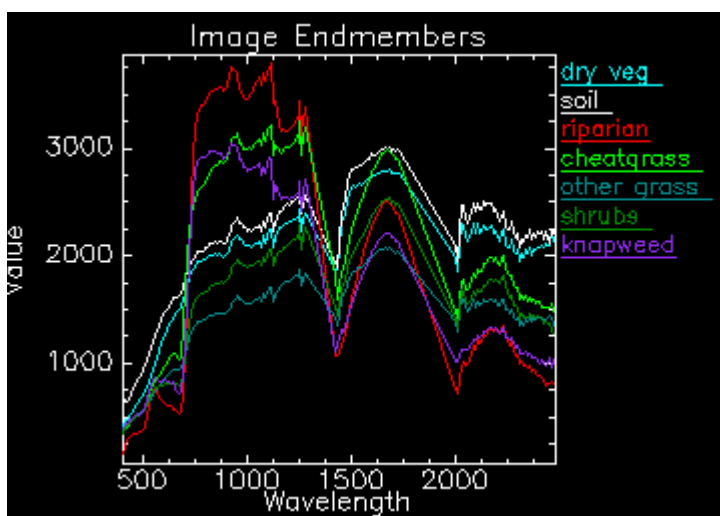

Figure 7-2: Comparison of image endmembers used for spectral mixture analysis of Yakima imagery.

ENVI linear spectral unmixing has two constraint options: unconstrained or a partially constrained unmixing. In unconstrained unmixing, abundances may assume negative values and are not constrained to sum to unity (one). An optional, variable-weight, unit-sum constraint in the Linear Mixing algorithm allows users to define weighting of a sum-to-unity constraint on the abundance fractions. It also permits proper unmixing of MNF-transform data, with zero-mean bands. The user picks a weight factor (default value = 1), for the extra constraint equation. This weighted unit-sum constraint is then added to the system of simultaneous equations in the unmixing inversion process. Larger weights in relation to the variance of the data cause the unmixing to reflect the unit-sum constraint more closely. To accurately represent the constraint, the weight is set to be many times the spectral variance of the data.

Lesson 10: Linear Spectral Unmixing

Required Files: `Ykm1a_reflectance.img`
                         `endmembers_reflectance.txt`

Generated Files: `Ykm1a_LSU.img`

<u>Applying Linear Spectral Unmixing</u>

1. Select **File → Open Image File →** `Ykm1a_reflectance.img`.

2. **Spectral → Mapping Methods → Linear Spectral Unmixing**. Select the `Ykm1a_reflectance.img` and click **OK**.

3. In the **Endmember Collection: Unmixing** dialog, **Import → From ASCII File →** `endmembers_reflectance.txt`, and click **OK**.

4. In the **Endmember Collection: Unmixing** dialog, select endmember spectra from `endmembers_reflectance.txt`. Click **Apply.** The Unmixing Parameters dialog appears.

5. Select **Yes** to "Apply a unit sum constraint" and use default weight of 1.0.

6. Name output `Ykm1a_LSU.img`.

7. Click **OK**.

<u>Spectral Unmixing Results</u>

The results of spectral unmixing appear as a series of gray-scale images, one for each endmember, plus a root-mean-square (RMS) error image. Higher abundances (and higher errors for the RMS error image) are represented by brighter pixels (larger floating-point numbers).

Lesson 11: Pixel Purity Index™ (PPI™) (Optional)

Required Files: `susbet_MNF.img`

Use **Pixel Purity Index** to find the most spectrally pure (extreme) pixels in multispectral and hyperspectral images. The most spectrally pure pixels typically correspond to mixing endmembers.

The Pixel Purity Index is computed by repeatedly projecting n-dimensional scatterplots onto a random unit vector. The extreme pixels in each projection—those pixels that fall onto the ends of the unit vector—are recorded and the total number of times each pixel is marked as extreme is noted. A Pixel Purity Image is created in which the DN of each pixel corresponds to the number of times that pixel was recorded as extreme.

The Pixel Purity Index function can create a new output band or continue its iterations and add the results to an existing output band. The PPI is typically run on a Minimum Noise Fraction (MNF) transform result excluding the noise bands. The results of the PPI are usually used as input into ENVI's n-D Visualizer.

FAST Pixel Purity Index

You have the choice of a disk-based PPI method and a FAST PPI method. The FAST PPI places the image data into memory and performs the computations in memory, which is much faster than the disk-based PPI method, but requires adequate memory space. Use spatial and spectral subsetting or a sparser array of the data (subsampled resolution) to help fit the image data into memory. ENVI informs you of the amount of memory necessary to perform the FAST PPI and prompts you for continuation. When using FAST PPI, you also have the options of creating a new output file and of adding to an existing output band.

Selecting Input Files

1. Use **New Output Band** the first time you run the Pixel Purity Index. An output band containing the number of times each pixel was found to be extreme (pure) is created and used as input into ENVI's n-D visualizer.

2. Select **Spectral → Pixel Purity Index → New Output Band** or **FAST New Output Band**.

3. When the Pixel Purity Index Input File dialog appears, open the file `subset_MNF.img`.
    Typically, run the PPI on an MNF transform result, and use spectral subsetting to exclude noise bands based on the eigenimages and eigenvalue plot. The MNF image has already been spectrally reduced to 15 bands.

4. Click **OK** to continue. The Pixel Purity Index Parameters dialog appears.

Setting the Number of Iterations

In the Pixel Purity Index Parameters dialog, designate the number of times the data will be projected onto the random vector by entering a number into the Number of Iterations text box.

The more iterations run, the better job ENVI does of finding the extreme pixels. Balance the number of iterations against the time available. [Each iteration can take some time depending on the CPU and system load.] Typically thousands of iterations are required for imaging hyperspectral data. The number of iterations run is listed in the description line of the image header file. The default value is 10,000 which should only take a few minutes on a small image.

Setting the Threshold Factor

In the **Threshold Factor** text box, enter a threshold value in data units for extreme pixel selection.

For example, a threshold of 2 marks all pixels greater than two digital numbers (DN) from the extreme pixels (both high and low) as being extreme. This threshold selects the pixels on the ends of the projected vector.

The threshold should be approximately 2-3 times the noise level in the data. For example, for Landsat TM data (which typically has less than 1 DN noise) a threshold of 2 or 3 works well. When using MNF data, which normalizes the noise, a DN is equivalent to 1 standard deviation and, again, a threshold value of 2 or 3 works well. Larger thresholds cause the PPI to find more extreme pixels, but they are less likely to be pure endmembers.

Subsampling the Data

Subsample the data to help it fit in memory. Enter X and Y resize factors less than 1 (eg. a resize factor of .5 will use every other pixel). It is not recommended to subsample by less than .25 (every fourth pixel) because extreme pixels may be thrown out.

Starting the Processing

If you are using FAST PPI, a window indicating the amount of memory needed appears and prompts you to continue if that amount of memory is acceptable.

1. In the Pixel Purity Index Parameters dialog, select **File** or **Memory** output.

2. Click **OK** to start the processing.

3. To interrupt PPI processing, click **Cancel** in the processing status window.

A processing status window appears with the Pixel Purity Index plot. This plot shows the total number of extreme pixels satisfying the threshold criterion found by the PPI processing as a function of the number of iterations. It should asymptotically approach a flat line (zero slope) when all of the extreme pixels have been found.

If it is difficult to visually distinguish pixel value differences, click on **Enhance → [Scroll] Equalization** within the main image window. This tool does not change the data, it merely enhances the visual contrast of the pixels.

Restarting or Continuing PPI

Use **Existing Output Band** or **[FAST] Existing Output Band** if you clicked Cancel and PPI was interrupted during processing, or use either option if you want to continue additional iterations on a PPI result.

1. Select **Spectral → Pixel Purity Index → Existing Output Band** or **[FAST] Existing Output Band**.

2. Select an input file (a spectrally subsetted MNF is recommended).

3. Click OK.

4. When the Pixel Purity Index Previous Result dialog appears, select a previous PPI image as the input file and click **OK**.

5. When the Pixel Purity Index Parameters dialog appears, select the number of iterations and threshold as described in the previous sections.

6. For FAST PPI, enter **X and Y Resize Factors** if desired.

7. Click **OK** to start the processing.

A processing status window and the Pixel Purity Index plot appear. The PPI image appears in the Available Bands List when the processing is completed.
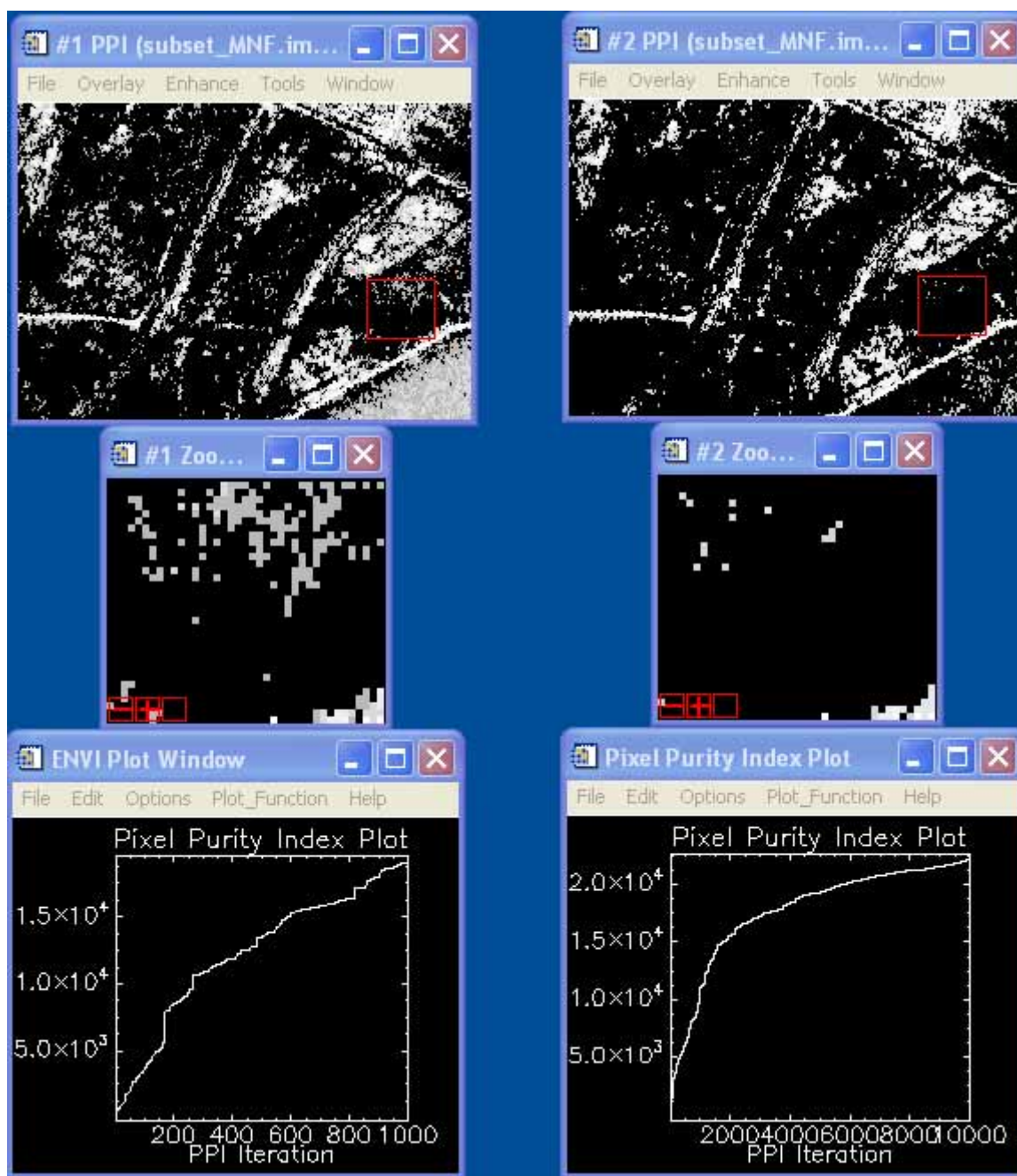
Figure 7-3: PPI results with 2.5 threshold, 1000 iterations (left) and 10,000 iterations (right)

Using PPI Images for Endmember Selection

After the PPI is processed, the PPI image appears in the Available Bands List. Display the image using standardized ENVI display procedures. Brighter pixels represent more

spectrally extreme hits and indicate pixels that are more spectrally pure. Darker pixels are less spectrally pure.

For the following steps, please review the techniques described in the ENVI Online Help.

8. Select **Window → Cursor Location/Value** on the ENVI main menu or **Tools → Cursor Location/Value** from the Display menu to determine the range of values present in the image.

9. Use interactive density slicing to get a better visual representation of the high values. In the image window, select **Tools → Color Mapping → Density Slice**.

10. In the image window, select **Overlay → Regions of Interest → Option → Band Threshold to ROI** to create an ROI containing only the pixels with high PPI values. Select the pixel purity file.

Typically, use a minimum threshold. For example, a minimum of 10 will include all of those pixels with PPI values greater than 10 in the ROI. However, if bad data points exist in the PPI image, both a minimum and maximum threshold can be used.

After an ROI has been created containing the high PPI values, you can use the **n-Dimensional Visualizer** to interactively define the image endmembers.

11. In the ENVI main menu, **Spectral → n-Dimensional Visualizer → Visualize with New Data**.

7.4   Mixture-Tuned Match Filtering (MTMF)

This section is adapted from the ENVI tutorial manual *(Research Systems, 2002)*

Mixture-Tuned Matched Filtering™ (MTMF™) is a hybrid method based on the combination of matched filtering and linear mixture theory (Boardman, 1998). This method combines the strength of the matched filter method with physical constraints imposed by mixing theory.

Matched filtering is a technique adapted from the electrical engineering field which maximizes the target-to-background contrast. A matched filter image is a result of maximizing the response of a known endmember and suppressing the response of the composite unknown background, thus matching the known signature (Chen and Reed, 1987; Stocker et al., 1990; Yu et al., 1993; Harsanyi and Chang, 1994). It provides a rapid means of detecting target spectra based on matches to specific library or image endmember spectra. This technique produces images similar to unmixing, but does not require the spectra of all endmembers. It does, however, suffer from high false alarm rates, where materials may be randomly matched if they are rare in a pixel (thus not contributing to the background covariance).

The equation for the MF image is as follows:

$$\Gamma = \frac{x^T \sum^{-1} d}{d^T \sum^{-1} d}$$

Where:
d = endmember spectrum
x = target spectrum
$\sum$ = covariance matrix

Mixture tuning uses linear spectral mixing theory to constrain the result to feasible mixtures and reduce false alarm rates (Boardman, 1998). MTMF results are presented as two sets of images: 1) the MF score (matched filter image presented as gray-scale images with values from 0 to 1.0) which provides a means of estimating relative degree of match to the reference spectrum (where 1.0 is a perfect match); 2) the infeasibility image indicates that mixing between the composite background and the target is not feasible. The best match to a target is obtained when the MF score is high (near 1) and the infeasibility score is low (near 0).

Lesson 12: Mixture-Tuned Matched Filtering

CD-ROM: *SERDP Training & Data CD*

Required Files: Ykm1b_MNF.img
                                 Ykm1b_mask.img
                                 Ykm1b_mnf_signature_knap.txt
                                 Ykm1b_mnf_signature_other.txt
                                 Ykm1b_reflectance.img

Generated Files: Ykm1b_MTMF.img
                                 Ykm1b_MTMF_sta.sta

Display MNF and reflectance Data

1. Select **File → Open Image File** and choose Ykm1b_MNF.img.

2. In the **Available Bands List**, click **RGB**, then choose MNF bands 4, 3, and 2.

3. Click **Load RGB**.

4. Select **File → Open Image File** and choose Ykm1b_reflectance.img.

5. In the **Available Bands List**, click **RGB** and choose Band 54 (red), Band 33 (green) and Band 19 (blue).

6. **Display → New Display → Load RGB**.

7. Link the two images using **Tools → Link → Link Displays**. Compare the MNF and reflectance images.

8. **Tools → Profiles → Z Profile (Spectrum)**. To observe the change in spectral profiles, move the cursor in either display.

Run MTMF

9. From ENVI's main menu, **Spectral → Mapping Methods → Mixture Tuned Matched Filter →** Ykm1b_MNF.img. Spectrally subset using Bands 2-24.

10. Select mask: Ykm1b_mask.img

11. From the **Endmember Collection** dialog, **Import → From ASCII →** Ykm1b_mnf_signature_knap.txt and Ykm1b_mnf_signature_other.txt.

12. Save statistics file as `Ykm1b_MTMF_sta.sta`. Save image file as `Ykm1b_MTMF.img`.

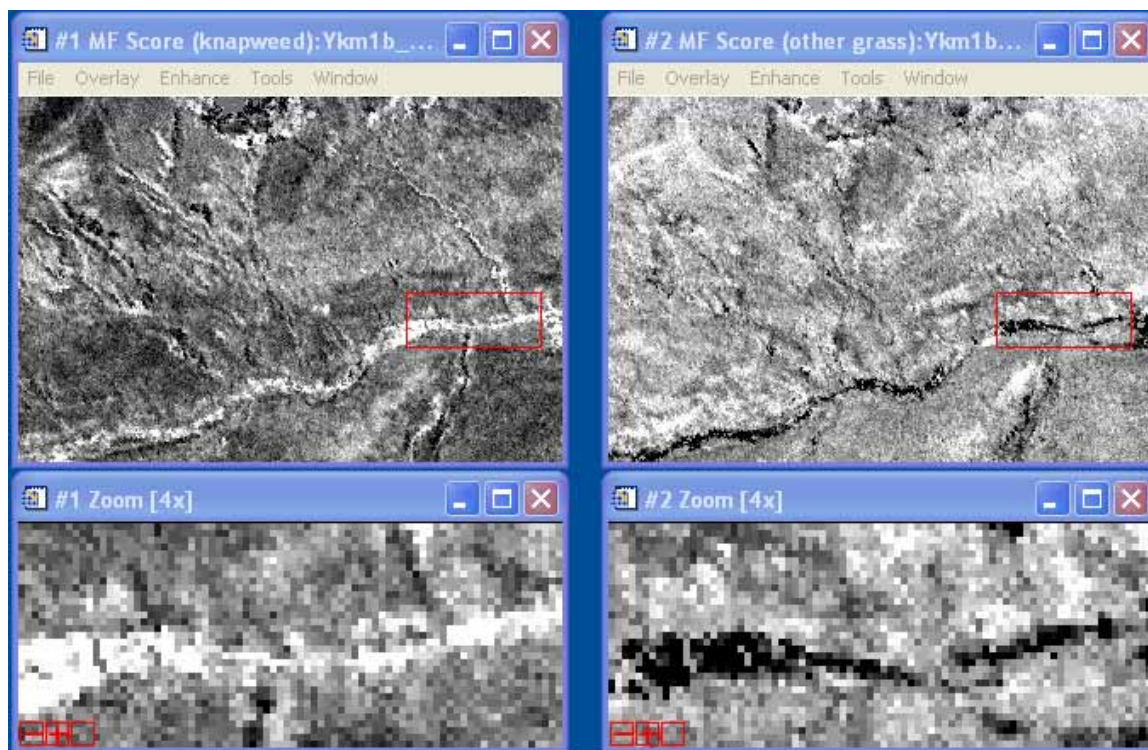13. Click **OK**. Compare the results for the two endmembers.

Figure 7-4: MTMF results for Russian knapweed (left) and "other vegetation" class (right).

8   Verification of Classification

After classifying the hyperspectral images using one or more of the techniques listed in Sections 5 through 7, a statistical analysis of the map accuracy is done with a subsequent fine-tuning of the classification as necessary.

Start by using some of the data collected during the preliminary fieldwork stage (Section 2), as test data for the classification verification. This data set contains plots and ROIs that were not used as part of the training set for the classifier, but were intentionally set aside to be used during the verification. The test data are ideally collected at the time of the overflight.

This first approximation of accuracy is used primarily to determine if the classifier should be more finely tuned prior to post-classification fieldwork. If the classifier is in reasonable agreement with the existing field data, then additional independent field validation is merited. If the classification results are not in reasonable agreement, the parameters are readjusted and the data is reprocessed to produce a new map. There are no hard rules for what can be considered reasonable, but accuracies above 50% are expected (at least for presence/absence of target invasive plant species).

8.1   Field Data Collection

The post-classification field measurements are usually a combination of stratified, semi-random, and user-selected sampling of ROIs.

The goal of sample site selection is to collect ROIs that focus on the areas of interest in invasive plant mapping, but are also geographically representative of the study area. Since field data collection is expensive and time-consuming, techniques for maximizing the data content for the minimum field effort are used, such as stratified sampling and boundary transects.

Stratified random sampling is a method of maximizing useful information by grouping land cover category into stratums and then randomizing the field sampling within each stratum (Lillesand and Kiefer, 2000). Semi-random sampling occurs when points are randomly selected across the area (e.g., the grid shown in Figure 8-1).
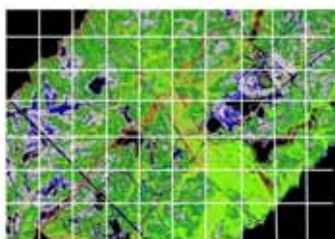

Figure 8-1: Semi-random block sampling within a field to distribute samples across the area. ROI are randomly selected within grid points.

As a broad guideline, Lillesand and Kiefer (2000) suggest that a minimum of 50 samples of each vegetation or land use category be acquired to perform the error statistics. A higher number of sampling points are placed in land covers of particular interest. In areas of low interest, further stratify the sampling area into topographic or some other ecological units to use as a mask and therefore, limit sampling.

In the case of invasive plant species mapping, the landscape factors that control the successful establishment of the species (e.g., disturbance, fire history, erosion, etc.) are important to consider in planning the strata. Also the method of dispersal if known (e.g., wind, animal vector or stream transport), is taken into consideration in siting the samples.

Prior to going into the field, review the random sampling points to determine if they will be accessible to the field crew by reviewing topographic data and other GIS layers. Inaccessible points are eliminated. Also review points for vulnerability to map registration error. Sampling points that fall close to the boundaries between classes may be eliminated and/or possibly other points are substituted that are several pixels away from the classification boundaries.

Besides the semi-random sampling described above, also create a user-selected sampling that focuses on specific areas of interest within the classification image. One area of interest is the boundary between different classes. One useful method is to place transects across predicted boundaries between native vegetation and invasive plant species to determine if the boundaries are correctly located in the imagery. The transect method is less prone to map registration errors than the point method.

ROIs are also placed in areas of disagreement between the various classification maps. The end result of the classification procedure is a series of classified maps, one from each classification technique. Overlay these maps in the GIS and identify areas of agreement and areas of disagreement. Now place sampling points and polygons of interest in both types of areas.

Further guidelines for sampling strategies are provided in Appendix A.

8.2   Confusion Matrices

A confusion matrix is a standard method of assessing accuracy by comparison between the classification and the ground truth information. The confusion matrix is calculated by comparing the location and class of each ground truth pixel with the corresponding location and class in the classification image. Each column of the confusion matrix represents a ground truth class, and the values in the column correspond to the classification image's labeling of the ground truth pixels in the classified image.

The number of pixels correctly classified are found along the diagonal of the confusion matrix table. The total number of pixels is the sum of all the pixels in all the ground truth

classes. The overall accuracy is the sum of the number of pixels correctly classified divided by the total number of pixels.

The kappa coefficient (k) is another measure of the accuracy of the classification. It is calculated by multiplying the total number of pixels in all the ground truth classes (N) by the sum of the confusion matrix diagonals (xkk), subtracting the sum of the ground truth pixels in a class times the sum of the classified pixels in that class summed over all classes (N), and dividing by the total number of pixels squared minus the sum of the ground truth pixels in that class times the sum of the classified pixels in that class summed over all classes.

Khat = (n * SUM Xkk) - SUM (Xk+ * X+i) / n2 - SUM (Xi+ * X+i)

where SUM = sum across all rows in matrix
Xi+ = marginal row total (row i)
X+i = marginal column total (column i)
n = # of observations

The kappa coefficient takes into account the off-diagonal elements of the contingency matrix (errors of omission and commission). Errors of commission represent pixels that belong to another class that are labeled as belonging to the class of interest. The errors of commission are shown in the rows of the confusion matrix. Errors of omission represent pixels that belong to the ground truth class but the classification technique has failed to classify them into the proper class. The errors of omission are shown in the columns of the confusion matrix.

The producer accuracy is a measure indicating the probability that the classifier has labeled an image pixel as Class A given that the ground truth is Class A (avoiding error of omission).

The user accuracy is a measure indicating the probability that a pixel is Class A given that the classifier has labeled the pixel as Class A (avoiding error of commission).

Lesson 13: Confusion Matrices

Required Files: `Ykm1c_mtmf.img`
                       `Ykm1c_ground_truth.roi`

Confusion Matrix Example

This example explains the values calculated for the confusion matrix, including the overall accuracy, kappa coefficient, confusion (contingency) matrix, errors of commission, errors of omission, producer accuracy, and user accuracy. ENVI calculates a confusion matrix using either a ground truth image or regions of interest (ROIs). Both produce an output similar to this example.

Confusion Matrix: {M6} (640x400x1)
Overall Accuracy = (131003/256000) 51.1730%
Kappa Coefficient = 0.2648

Ground Truth (Pixels)					
Class	**Unclassified**	**Grass**	**Forest**	**Swamp**	**Total**
Unclassified	43689	26949	40	18001	88679
Grass	32835	64516	1741	3329	102421
Forest	8202	7277	4096	654	20229
Swamp	15227	10742	0	18702	44671
Total	99953	109484	5877	40686	256000

Ground Truth (Percent)					
Class	**Unclassified**	**Grass**	**Forest**	**Swamp**	**Total**
Unclassified	43.71	24.61	0.68	44.24	34.64
Grass	32.85	58.93	29.62	8.18	40.01
Forest	8.21	6.65	69.70	1.61	7.90
Swamp	15.23	9.81	0.00	45.97	17.45
Total	100.00	100.00	100.00	100.00	100.00

Class	**Commission (Percent)**	**Omission (Percent)**	**Commission (Pixels)**	**Omission (Pixels)**
Unclassified	50.73	56.29	44990/88679	56264/99953
Grass	37.01	41.07	37905/102421	44968/10944
Forest	30.32	30.30	6133/20229	1781/5877
Swamp	58.13	54.03	25969/44671	21984/40686

Class	Prod. Acc. (Percent)	User Acc. (Percent)	Prod. Acc. (Pixels)	User Acc. (Pixels)
Unclassified	43.71	49.27	43689/99953	43689/88679
Grass	58.93	62.99	64516/109484	64516/102421
Forest	69.70	20.25	4096/5877	4096/20229
Swamp	45.97	41.87	18702/40686	18702/44671

Confusion Matrix (Pixels)

The confusion matrix is calculated by comparing the location and class of each ground truth pixel with the corresponding location and class in the classification image. Each column of the confusion matrix represents a ground truth class and the values in the column correspond to the classification image's labeling of the ground truth pixels. For example, look at the ground truth column for the Forest class in the Ground Truth (Pixels) table above. The ground truth shows 5,877 pixels in this class. The classification was able to classify 4,096 of these pixels properly but 40 pixels were Unclassified and 1,741 were classified as Grass.

Confusion Matrix (Percent)

The Ground Truth (Percent) table shows the class distribution in percent for each ground truth class. The values are calculated by dividing the pixel counts in each ground truth column by the total number of pixels in a given ground truth class. For example, in the Forest class the percent pixels classified correctly is 4,096/5,877=0.697 or 69.7%.

Commission

Errors of commission represent pixels that belong to another class that are labeled as belonging to the class of interest. The errors of commission are shown in the rows of the confusion matrix. In the confusion matrix example, the Grass class has a total of 102,421 pixels of which 64,516 pixels are classified correctly and 37,905 other pixels are classified incorrectly as Grass (37,905 is the sum of all the other classes in the Grass row of the confusion matrix). The ratio of the number of pixels classified incorrectly divided by the total number of pixels in the ground truth class forms an error of commission. For the Grass class the error of commission is 37,905/102,421 or 37%.

Omission

Errors of omission represent pixels that belong to the ground truth class, but the classification technique has failed to classify them into the proper class. The errors of omission are shown in the columns of the confusion matrix. In the confusion matrix example, the Grass class has a total of 109,484 ground truth pixels of which 64,516 pixels are classified correctly and 44,968 are classified incorrectly (44,968 is the sum of all the other classes in the Grass column of the confusion matrix). The ratio of the number of pixels classified incorrectly by the total number of pixels in the ground truth class forms

an error of omission. For the Grass class the error of omission is 44,968/109,484 or 41.1%.

Producer Accuracy

The producer accuracy measures the probability that the classifier has labelled an image pixel into Class A given that the ground truth is Class A. In the confusion matrix example, the Grass class has a total of 109,484 ground truth pixels of which 64,516 pixels are classified correctly. The producer accuracy is the ratio 64,516/109,484 or 58.9%.

User Accuracy

The user accuracy measures the probability that a pixel is Class A given that the classifier has labelled the pixel into Class A. In the confusion matrix example, the classifier has labelled 102,421 pixels as the Grass class and 64,516 of them are classified correctly. The user accuracy is the ratio 64,516/102,421 or 63.0%.

Accuracy Assessment for SERDP Image

ENVI calculates a confusion matrix (contingency matrix) using either a ground truth image or regions of interest (ROIs). Overall accuracy, producer and user accuracies, kappa coefficient, confusion matrix, and errors of commission and omission are reported.

Using Ground Truth Regions of Interest

1. **File → Open Image File →** `Ykm1c_mtmf.img`.

2. In **Display #1 → Overlay → Region of Interest.**

3. In **#1 ROI Tool → File → Restore ROIs →** `Ykm1c_ground_truth.roi`.

4. Select **Classification → Post Classification → Confusion Matrix → Using Ground Truth ROIs**.

5. In the **Classification Input File** dialog, select `Ykm1c_mtmf.img`. Click **OK.**

6. In the **Match Classes Parameters** dialog, match the ground truth ROIs with the classification results. Click on the matching names in the two lists. Do this for all pairs.

7. Click **Add Combination** and then **OK**.

8. In the Confusion Matrix Parameters dialog, select **Pixels** and **Percent**. Click **Yes** and then **OK**.

9. How accurate was your SERDP classification?

9   Literature Citations

Adams, J. B., Sabot, D.E., Kapos, V., Almeida Filho, R., Roberts,D.A., Smith, M.O., and Gillespie, A.R. (1995). Classification of multispectral images based on fractions of endmembers: application to land cover change in the Brazilian Amazon. *Remote Sensing of Environment*, 52, 137-154.

Anderson, L. W. (1987), Hydrilla (*Hydrilla verticillata*). Exotic Pest Profile No. 11. California Dept. of Food and Agriculture.

Boardman, J.W. and F.A. Kruse. 1994. Automated spectral analysis: a geologic example using AVIRIS data, north Grapevine Mountains, Nevada. In: *Proceedings, Tenth Thematic Conference on Geologic Remote Sensing.* Environmental Research Institute of Michigan, Ann Arbor, MI, pI-407 – I-408.

Boardman, J. W., Kruse, F. A., and Green, R. O. 1995. Mapping target signatures via partial unmixing of AVIRIS data. In: *Fifth JPL Airborne Earth Science Workshop,* JPL Publication 95-1, v. 1:23-26.

Boardman, Joseph, 1998. Leveraging the high dimensionality of AVIRIS data for improved sub-pixel target unmixing and rejection of false positives: mixure tuned matched filtering. In: *Summaries of the Seventh Annual JPL Airborne Geoscience Workshop*, Pasadena, CA.

Bossard, C. C., Randall, J. M., and Hoshovsky, M. C.,(Eds) 2000. *Invasive plants of California's wildlands*. University of California Press.

Card, D. H., Peterson, D.L, Matson, P.A. and J.D. Aber. 1988. "Prediction of leaf chemistry by the use of visible and near infrared reflectance spectroscopy." *Remote Sensing of Environment* 26: 123-147.

Carter, G. A. (1991). "Primary and secondary effects of water content on the spectral reflectance of leaves." *American Journal of Botany* 78(7): 916-924.

Center for the Study of Earth from Space (CSES). 1992. *SIPS User's Guide, The Spectral Image Processing System*, v.1.1, University of Colorado, Boulder, 74 p.

Chen, J.Y. and I.S. Reed. 1987. A detection algorithm for optical targets in clutter. *IEEE Transactions on Geoscience and Remote Sensing*, v. 26, no. 1, p. 65-74

Cocks, T., Jennsen, R., Stewart, A., Wilson, I. And Shields, T.(1998). *The AVIRIS airborn hyperspectral sensor: the system, calibration and performance*, Proceedings of the 1st EARSEL Workshop on Imaging Spectroscopy, Zurich, October 1998.

Corliss, J. 1993. Tall whitetops crowding out the natives. *Agricultural Research*. May 1993: 16. In Bossard, C. C., Randall, J. M., and Hoshovsky, M. C.,(Eds) 2000. *Invasive plants of California's wildlands*. University of California Press.

Curran, P. J. (1989). "Remote sensing of foliar chemistry." *Remote Sensing of Environment* 30: 271-278.

Elmore, A. J., Mustard, J. F., Manning, S. J., and Lobell, D. B. (2000). Quantifying vegetation change in semiarid environments: precision and accuracy of spectral mixture analysis and the normalized difference vegetation index. *Remote Sensing of Environment*, 73, 87-102.

Fourty, T. and F. Baret (1998). "On spectral estimates of fresh leaf biochemistry." *International Journal of Remote Sensing* 19(7): 1283-1297.

Fuentes, D. A., Gamon, J.A., Qui H.L., Sims, D.A. and D.A. Roberts. 2001. "Mapping Canadian boreal forest vegetation using pigment and water absorption features derived from the AVIRIS sensor." *Journal of Geophysical Research* 106(D24): 33565-33577.

Gao, B.C. (1996). "NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space." Remote Sensing of Environment 58: 257-266.

Green, A. A., Berman, M., Switzer, P. and Craig, M. D. 1988 A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transactions on Geoscience and Remote Sensing*, 26(1):65-74.

Harsanyi, J.C. and C.I. Chang. 1994. Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach. *IEEE Transactions on Geoscience and Remote Sensing,* 32:779-785

Hortwitz, H. M., Lewis, J. T., and Pentland, A. P. (1975). Estimating proportions of objects from Multispectral Scanner Data, Final Report. NASA-CR 141862. 108pp.

Jackson, R. D. and A. R. Huete (1991). "Interpreting vegetation indices." *Preventive Veterinary Medicine* 11: 185-200.

Kruse, F.A., Lefkoff, A.B., Boardman, J.W., Heidebrecht, K.B., Shapiro, A.T., Barloon, J.P., and A.F.H. Goetz. 1993. The spectral image processing system (SIPS) – Interactive visualization and analysis of imaging spectrometer data. *Remote Sensing of the Environment* 44:145-163

Langeland, K. (1990). Hydrilla: a continuing problem in Florida waters. Florida Cooperative Extension Circular No. 884. Miami, FL., In Bossard, C. C., Randall, J. M., and Hoshovsky, M. C.,(Eds) 2000.*Invasive plants of California's wildlands*. University of California Press

Lichtenthaler, H. K., Lang, M., Sowinska, M., Heisel, F. and J.A. Miehe. 1996. "Detection of vegetation stress via a new high resolution fluorescence imaging system." *Journal of Plant Physiology* 148(5): 599-612.

Lillesand, T.M. and R.W. Kiefer. 2000. *Remote Sensing and Image Interpretation.* New York, NY. John Wiley & Sons, Inc. 750pp.

Mooney, H. A., and Cleland, E. E. (2001). The evolutionary impact of invasive species. *Procedures of National Academy of Science*, 98 (10), 5446-5451.

Nagler, P. L., Daughtry, C.S.T., and S.N. Goward.   2000. "Plant litter and soil reflectance." *Remote Sensing of Environment 71*: 207-215.

Nelson, L. S. and Getsinger, K. D. (1994). Battling purple loosestrife brings some success. The Wetlands Research Progrm Bulletin, Vol. 4, No. 1, pp. 5-8. Washington DC. In Bossard, C. C., Randall, J. M., and Hoshovsky, M. C.,(Eds) 2000. *Invasive plants of California's wildlands*. University of California Press

Parson, J. M. (1992). (ed). Australian Weed Control Handbook, Inkata Press, Melbourne, Australia. In Bossard, C. C., Randall, J. M., and Hoshovsky, M. C.,(Eds) 2000. *Invasive plants of California's wildlands*. University of California Press

Penfound, W. and Earle, T., (1948). The bilogy of the water hyacinth. Ecological Monographs. 18: 447-72. In Bossard, C. C., Randall, J. M., and Hoshovsky, M. C.,(Eds) 2000. *Invasive plants of California's wildlands*. University California Press

Penuelas, J., Baret, F. and I Filella. 1995. "Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance." *Photosynthetica* 31(2): 221-230.

Penuelas, J., Pinol, J., Ogaya, R., and I. Filella. 1997. "Estimation of plant water concentration by the reflectance Water Index WI (R900/R970)." *International Journal of Remote Sensing* 18(13): 2869-2875.

Pimm, S.L.., and Gilpin, M.E. (1989). Theoretical issues in conservation biology. In: Roughgarden, J., R., May and S.A. Levin (eds). Perspectives in Ecological Theory. Princeton University Press, Princeton, NJ. Pp 287-205.

Rahman, A. F., Gamon, J.A., Fuentes D.A., Roberts D.A. and D. Prentiss. 2001. "Modeling spatially distributed ecosystem flux of boreal forest using hyperspectral indices from AVIRIS imagery." *Journal of Geophysical Research* 106(D24): 33579-33591.

Research Systems. 2002. *ENVI Tutorials.* December 2002 Edition. Research Systems, Inc. A Kodak Company,

Riano, D. Chuvieco, E. , Ustin, S. , Zomer, R. , Dennison, P., Roberst, D. , and Salas, J. (2002). Assessment of vegetation regeneration after fire through multitemporal analysis of AVIRIS images in the Santa Monica Mountains. *Remote Sensing of Environment*, 79, 60-71.

Roberts, D.A., Green, R. AO., and Adams, J. B. (1997). Temporal and spatial patterns in vegetation and atmospheric properties from AVIRIS. *Remote Sensing of Environment*, 62, 223-240.

Sabol, D. E., Adams, J. B. and Smith, M. O. 1992. Quatitative subpixel spectral detection of targets in Multispectral Images. J. Geophys. Res., 97(E2): 2659-2672.

Sanderson, E.W., M. Zhang, S.L. Ustin, and E. Rejmankova. 1998. Geostatistical scaling canopy water content in a California salt marsh. Landscape Ecology 13:79-92.

Serrano, L., Penuelas, J, and S.L. Ustin. 2002. "Remote sensing of nitrogen and lignin in Mediterranean vegetation from AVIRIS data: decomposing biochemical from structural signals." *Remote Sensing of Environment* 81: 355-364.

Smith, M. O., Ustin, S. L., Adams, J. B., and Gillespie, A. R. (1990). Vegetation in deserts: I. A regional measure of abundance from multispectral images. *Remote Sensing of Environment*, 31, 1-26.

Stocker, A.D., Reed, I.S. and X.Yu. 1990. Multidimensional signal processing for electrooptical target detection. *Proceedings, SPIE International Society of Optical Engineers.* Vol. 1305.

Underwood, E., S. Ustin, and D. DiPietro. 2003. Mapping nonnative plants using hyperspectral imagery. *Remote Sensing of Environment* 86:150-161.

Ustin, S.L., M.O. Smith, and J.B. Adams. 1993. Remote Sensing of Ecological Processes: A strategy for Developing Ecological Models Using Spectral Mixture Analysis.   In J. Ehlringer and C. Field (Eds.) *Scaling Physiological Processes: Leaf to Globe*. Academic Press, New York, p.339-357.

Ustin, S.L., D. DiPietro, K. Olmstead, E. Underwood, and G.J. Scheer. 2002. Hyperspectral remote sensing for invasive species detection and mapping. In: *Proceeedings of International Geoscience and Remote Sensing Symposium,* IGARSS-2001. Toronto, Canada. June 24-28, 2002

Ustin, S. L. and S. Jacquemoud. 2002. "Leaf optical properties." *Digital Photobiology Compendium*. www.photobiology.info

Ustin, S.L., D.A. Roberts, S. Jacquemoud, J. Pinzon, M. Gardner, G. Scheer, C.M. Castaneda, and A. Palacios. 1998. Estimating canopy water content of chaparral shrubs using optical methods. *Remote Sensing of Environment* 65:280-291.

Yu, X., Reed, I.S., and A.D. Stocker. 1993. Comparative performance analysis of adaptive multispectral detectors. *IEEE Transactions on Signal Processing* 41(8).

Zarco-Tejada, P. J. (1998). Optical indices as bioindicators of forest sustainability. Graduate Programme in Earth and Space Science. Toronto, York University.

Zhang, M., S.L. Ustin, E. Rejmankova, and E.W. Sanderson. 1996. Remote sensing of salt marshes: Potential for monitoring. Ecological Applications 7(3): 1039-1053.

Appendix A: Sampling Strategies

The following material is derived from the website of Offwell Woodland & Wildlife Trust is copyrighted, but is copied with permission (2003) of the nonprofit organization for British wildlife conservation and environmental education.
http://www.countrysideinfo.co.uk/index.htm

The problem of characterizing a habitat is usually solved by taking a number of samples from around the habitat, making the necessary assumption that these samples are representative of the habitat in general. In order to be reasonably sure that the results from the samples do represent the habitat as closely as possible, careful planning beforehand is essential.

Usually samples are taken using a standard sampling unit so that all of them represent the same area or volume (water) of the habitat.

Quadrats

The quadrat, a typical sampling unit, usually consists of a square frame of 1 m^2 (see picture below). Using quadrats brings a consistent size and shape to the areas sampled. Rectangular and even circular quadrats have been used in some surveys. A quadrat's shape does not matter as much as its being a uniform sampling unit and its shape and measurements are stated in any write-up.
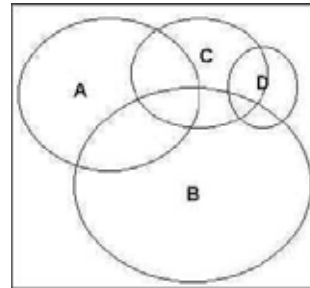
A quadrat is used to sample the species present in a habitat.

Choice of quadrat size depends to a large extent on the type of survey being conducted. As a general guideline, 0.5 - 1.0 m^2 quadrats are suggested for short grassland or dwarf heath. Taller grasslands and shrubby habitats might require 2m quadrats and woodland habitats would need quadrats of 20m^2 or larger.

To record percent cover of species in a quadrat, look down on the quadrat from above and estimate the percent cover occupied by each species (e.g. species A - D *left*). Species often overlap and there may be several different vertical layers. Percent cover may therefore add up to well over 100% for an individual quadrat.

The estimation can be improved by dividing the quadrat into a grid of 100 squares - each representing 1% cover. Either visualize 10 longitudinal and 10 horizontal lines of equal size superimposed on the quadrat, or physically divide the quadrat with string or wire attached to the frame at standard intervals.
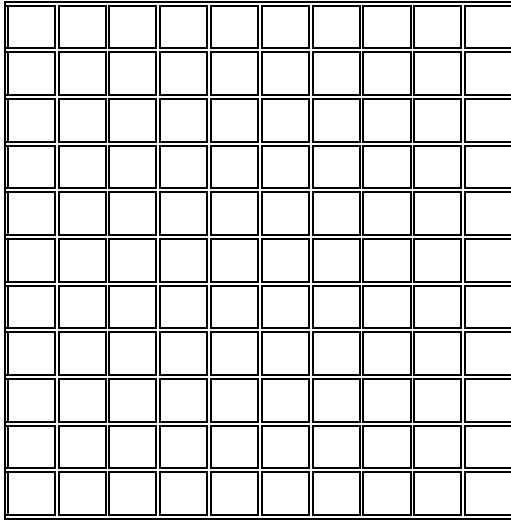
Quadrats are most often used for sampling, but are not the only type of sampling units. It depends on what you are sampling. If you are sampling aquatic microorganisms or studying water chemistry, then most likely you will collect water samples in standard sized bottles or containers. If you are looking at parasites on fish, then an individual fish probably will be your sampling unit. Similarly, studies of leaf miners would involve collecting individual leaves as sampling units. In these last two cases, the sampling units will not be of standard size. This problem can be overcome by using a weighted mean, which takes into account different sizes of sampling unit, to arrive at the mean number of organisms per sampling unit.

Random Sampling

Random sampling is usually carried out when the area under study is fairly uniform, very large, and/or there is limited time available. When using random sampling techniques, large numbers of samples/records are taken from different positions within the habitat. A quadrat frame is most often used for this type of sampling. The frame is placed on the ground (or on whatever is being investigated) and the animals, and/ or plants inside it are counted, measured, or collected, depending on the intent

of the survey. Sampling is repeated many times at different points within the habitat to obtain a fair representation of the study site.

In the simplest form of random sampling, the quadrat is thrown at random within the site. Many common statistical techniques are valid only on data that are randomly collected. This technique is possible only if quadrats of small size are used.

An alternative method of random sampling is to map the area and then to lay a numbered grid over the map. A (computer generated) random number table is then used to select in which squares to sample. For example, if our habitat is mapped from the hyperspectral data and a numbered grid is layed over it as shown the Figure below, then locations of sampling points can be chosen using the random number table.
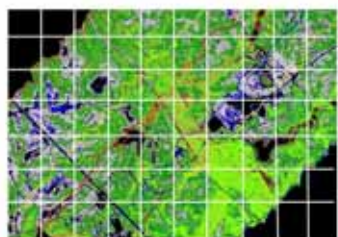
In some habitats such as this woodland, it is difficult to set up numbered grids and in these a random walk may be used. In this method, each sample point is located by selecting a random number between 0 and 360, to give a compass bearing, followed by another random number which indicates the number of paces which should be taken in that direction. A grid can be established in advance to ensure sampling occurs throughout the area.

Line Transect Method

A transect line can be made using a nylon rope marked and numbered at 0.5 m, or 1 m intervals. This line is laid across the area to be surveyed. Locating the position of the transect line is critical to the results and it depends on the direction of the environmental gradient. A line transect is conducted by unrolling the transect line along the gradient identified. The species touching the line may be recorded along the whole length of the line (continuous sampling). Alternatively, the presence, or absence of species at each marked point is recorded (systematic sampling). If the slope along the transect line is measured as well, the results can then be inserted onto a profile of the transect.

Belt Transect Method

This method is similar to the line transect, but gives information on abundance as well as the presence or absence of a species. It may be considered as a widening of the line transect to form a continuous belt, or series of quadrats.
In this method, the transect line is laid across the area to be surveyed and a quadrat is placed on the first marked point on the line. The plants and/or animals inside the

quadrat are then identified and their abundance estimated and the percent cover of plant species.

Quadrats can be sampled the length of the transect line (a belt transect), at each marked point on the line, or at some other predetermined interval (or even randomly). It is best to have the same person estimate cover because it is likely to vary from person to person. The height of plants can be recorded and the biomass of plants can also be measured by harvesting all the plants inside the quadrat and then weighing either fresh, or dry weight in the laboratory. Sampling should always be as least destructive as possible and you should try not to trample an area too much when carrying out your survey.

An example of the type of results that can be obtained from a belt transect survey is shown below. This figure illustrates the distribution and abundance of cherry seedlings along a transect line. The parent cherry trees were adjacent to section number 9. The gradient of distribution apparent in the figure is a result of the dispersal of seeds outwards from this point.
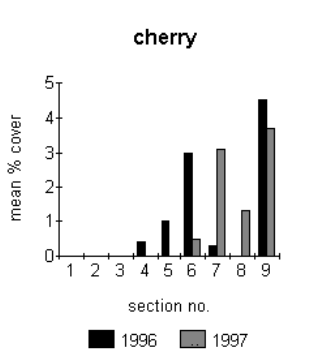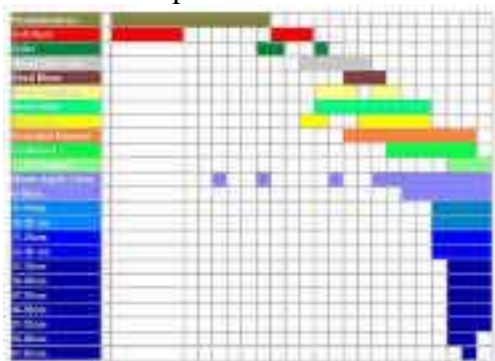
Stratified Sampling
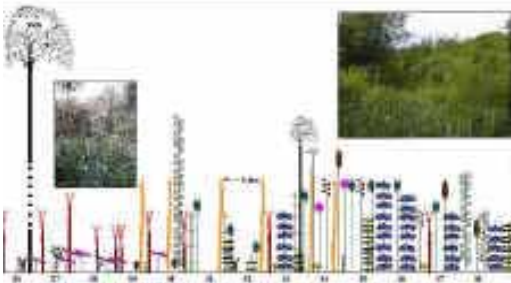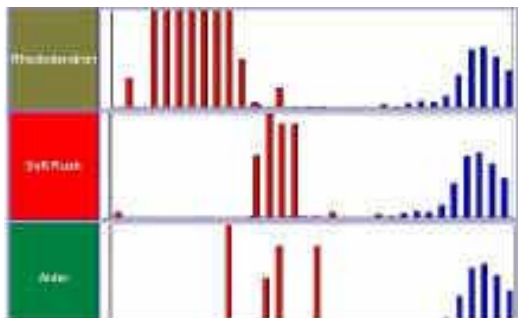
Stratified sampling is used where there are small areas within a larger habitat which are clearly different. For example, scrub patches within a heathland area, or areas of bracken in a grassland. Sampling would be carried out either randomly, or systematically within each separate stratum identified. Major differences within communities are recognized before sampling begins.

Choosing a Sampling Interval

Samples can be taken continuously or at specific points along a transect. For both line and belt transects, the interval at which samples are taken depends on the individual habitat, as well as on the time and effort which can be allocated to the survey.

- Too large an interval – species present may be missed and zonation patterns obscured.

- Too small an interval – too time consuming and produces more data than needed.

- Coincident with regularly occurring landscape features – avoid. For instance, you do not want all samples on ridges in an old field with ridge and furrow systems.

- Ideal interval - balances the complexity of the individual habitat with the purpose of the survey and the resources available to carry it out.

Decision Tree for Selecting Sampling Method

Random Sampling?	**Systematic Sampling (Transects)?**
This is used where the habitat being sampled is fairly uniform. To remove observer bias in the selection of samples. Where statistical tests are to be used which require randomly collected data. Where a large area needs to be covered quickly. If time is very limited. **How many samples?**	To show zonation of species along some environmental gradient. e.g. across an edge between community types. Where there is some kind of continuous variation along a line, To sample linear habitats, e.g. a roadside habitat or riparian zone. Where physical conditions demand it, e.g. sampling a vertical rock face.

Systematic Sampling - Line or Belt?

⬇	⬇
<u>Line Transect?</u>	**Belt Transect?**

Where time is limited. A line transect can be measured quicker than a belt transect. To visually illustrate how species change along the line. Keys can be chosen to represent individual species. Vegetation height can be drawn in choosing an appropriate scale. The slope of the line can also be measured when measuring the transect and then incorporated into the transect diagram to show species distributions along the line. .	A belt transect will supply more data than a line transect. It will give data on the abundance of individual species at different points along the line, as well as on their range. As well as showing species ranges along the line, a belt transect will also allow bar charts to be constructed showing how the abundance of each individual species changes within its range.
This method generally shows only where the species occurs, not how much of it is present.	Belt transect data allow the relative dominance of species along the line to be determined.